Supervised control, trust and dialogue

Cyril Crocquesel

N° d'ordre : 2012telb0232

Sous le sceau de l'Université européenne de Bretagne

Télécom Bretagne

En habilitation conjointe avec l'Université de Bretagne Sud

École Doctorale – SICMA

Contrôle supervisé, confiance et dialogue

Thèse de Doctorat

Mention : « Sciences et technologies de l'information et de la communication »

Présentée par **M Cyril Crocquesel**

Département : LUSSI

Laboratoire : Lab-STICC - Pôle CID

Directeur de thèse : M. Gilles Coppin, Professeur à Télécom Bretagne

Soutenue le 27 septembre 2012

Jury :

Rapporteurs :	M. Frédéric Vanderhaegen, Professeur à l'université de Valenciennes
	M. Frédéric Dehais, Professeur à l'ISAE
Présidente :	Mme Christine Chauvin, Professeur à l'université de Bretagne Sud
Examinateur :	M. Jean-Yves Antoine, Professeur à l'université Francis Rabelais (Tours)
Encadrant :	M. François Legras, Docteur ingénieur à Deev Interaction

Remerciements

Le point final de cette thèse sera posé à la fin de ces quelques lignes. Bien qu'elles figurent en début de mémoire, elles en sont pourtant les derniers mots écrits.

Je remercie Gilles Coppin, directeur de thèse, pour son suivi et l'aide constante qu'il m'a apporté dans la réalisation de la thèse et du mémoire.

Je remercie François Legras, encadrant de thèse, pour les conseils qu'il m'a apporté et le temps qu'il aura consacré à la correction de ce mémoire.

Je remercie Brest Métropole Océane qui a financé les travaux de recherches de cette thèse (Contrat de recherche PAGICEA 3CCPT813 : 2008-2011).

Je remercie le département LUSSI et tous ses membres pour leur accueil et leur soutien tout au long la thèse.

Je remercie ma famille et mes amis pour leurs encouragements et leur aide.

Enfin je remercie tout ceux que j'ai sûrement oublié ici mais qui ont tout autant droit à ma gratitude, de par leurs conseils, leur soutien et leur participation.

Table des matières

Introduction

Qu'y a-t-il de commun entre une centrale nucléaire, un réseau de distribution des télécommunications, le contrôle du trafic ferroviaire, un avion de ligne ou le contrôle de drones ? Ce sont tous des systèmes qui comportent un grand nombre d'automates. Ces derniers nécessitent une supervision humaine. Ainsi, une centrale nucléaire est surveillée par une équipe d'opérateurs, le trafic ferroviaire par ses aiguilleurs, un avion par son équipage et les drones par leurs opérateurs. Chacun de ces êtres humains interagit avec ces automates, et leur porte une confiance plus ou moins importante. C'est pour cette raison, qu'au cours d'un vol de la Eastern Air Lines, les membres de l'équipage ne prêtaient pas attention au pilote automatique et n'ont pas détecté son désenclenchement. Les conséquences directes ont été une perte d'altitude de l'appareil et le déclenchement d'une alarme qui indique un écart entre la consigne d'altitude et l'altitude réelle de l'avion. Cette alarme a été ignorée par l'équipage. Ce n'est que quelques minutes plus tard que le pilote, contacté par la tour de contrôle, prend conscience de la perte d'altitude. Il est alors trop tard pour une partie de l'équipage et des passagers qui périrent lors du crash de l'appareil. La confiance trop importante de l'équipage les a conduit à une erreur de supervision du pilote automatique.[1]

L'excès ou le manque de confiance des opérateurs peut conduire à deux problèmes : les usages inappropriés de l'automate (*Misuse*) et le non usage de l'automate (*Disuse*). En effet les capacités peuvent être mal appréciées par un opérateur. Ainsi une sous-estimation (ou une surestimation) des capacités de l'automate, ou une confiance inadaptée, trop faible (ou trop forte), de la part de l'opérateur engendreront un usage inapproprié de l'automate. Dans d'autres cas, l'opérateur a un tel manque de confiance qu'il abandonne l'usage de l'automate au profit d'opérations manuelles, ce qui correspond au non usage de l'automate.

Or, si des systèmes automatiques et robotiques se substituent de plus en plus à l'homme, ce n'est pas pour qu'ils soient rejetés par leur opérateur et, de ce fait, que les risques encourus augmentent. Il est nécessaire d'entretenir au mieux une relation de confiance entre l'homme et la machine. Pour cela, il faut estimer le degré de confiance des opérateurs qui évolue selon leur expérience accumulée au cours de leur relation avec ces systèmes. Cette relation repose sur un échange d'informations entre l'opérateur et son système, c'est-à-dire sur l'interaction. C'est pourquoi nous proposerons dans ce mémoire une nouvelle approche pour l'évaluation de la confiance dans les interactions Homme-machine.

Nous proposons dans un premier temps un état de l'art, divisé en trois parties, portant sur :

[1]National Transportation Safety Board (1972). Eastern Airlines L-1011. Miami. Florida. 29 December 1972 (Rep. NTSB-AAR-73-14). Washington, DC.

– le contrôle supervisé (chapitre 1) afin de présenter le contexte de notre étude et les modèles qui lui sont rattachés ;
– la confiance (chapitre 2) qui est l'un des facteurs humains majeurs mis en jeu au sein du contrôle supervisé ;
– le dialogue (chapitre 3) dont l'analyse sera la base de notre évaluation.

Nous présenterons ensuite une étude exploratoire issue d'une première campagne expérimentale (chapitre 4). Puis nous introduirons notre approche de l'évaluation de la confiance (chapitre 5). Et enfin, afin de valider notre modèle d'évaluation, nous détaillerons une seconde campagne expérimentale (chapitre 6).

1 Contrôle supervisé

Les relations Hommes-machines ont longtemps reposé sur le paradigme du contrôle manuel. L'automatisation croissante des systèmes a nécessité son évolution pour aboutir au contrôle supervisé. Nous commencerons cet état de l'art par la description théorique du contrôle d'un automate par un opérateur. Nous étudierons dans un second temps les problèmes sous-jacents au contrôle supervisé.

1.1 Définition

On peut situer le contrôle supervisé au sein d'une hiérarchie allant du contrôle manuel au système entièrement automatique (illustrée en figure 1.1) :

- Contrôle manuel direct de l'opérateur sur la tâche ;
- Contrôle manuel de l'opérateur sur la tâche via un ordinateur ;
- Contrôle supervisé par un opérateur qui agit toujours sur la tâche via l'ordinateur mais l'ordinateur est autonome lors de la non-intervention de l'opérateur (déf. 1.1) ;
- Contrôle supervisé par un opérateur qui ne peut pas agir sur la tâche mais uniquement sur l'ordinateur. Ce dernier assure donc le contrôle complet de la tâche selon les directives de l'opérateur (déf. 1.2) ;
- Système complètement autonome, l'opérateur a uniquement un retour d'informations.

Sheridan définit alors le contrôle supervisé de deux façons [102], la première étant moins stricte que la seconde :

Déf. 1.1 (Contrôle supervisé) *Un (ou plusieurs) opérateur(s) humain(s) programme(nt) et reçoit (reçoivent) des informations de manière continue d'un ordinateur qui le(s) relie au contrôle d'une procédure et à l'environnement de la tâche via des capteurs et des effecteurs artificiels.*

Déf. 1.2 (Contrôle supervisé strict) *Un (ou plusieurs) opérateur(s) humain(s) programme(nt) ponctuellement un ordinateur et reçoit (reçoivent) en continu des informations de ce même ordinateur. Ce dernier possède une boucle de contrôle autonome pour contrôler la procédure et l'environnement de la tâche via des capteurs et des effecteurs artificiels.*

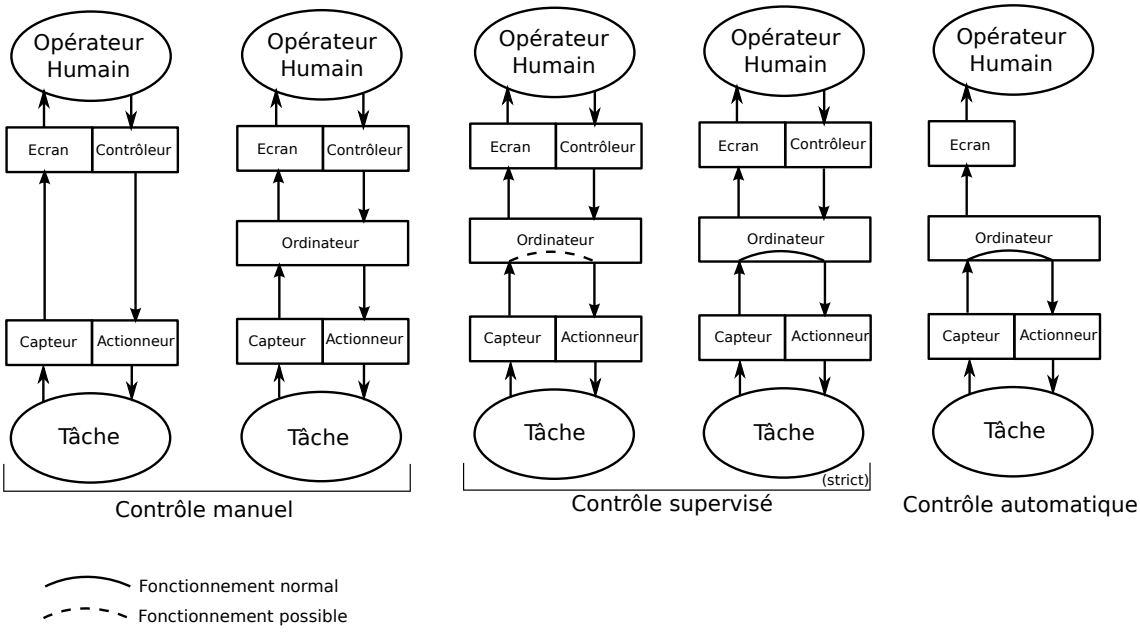

FIG. 1.1 : Hiérarchie des divers modes de contrôle - d'après Sheridan [102].

À l'aide de cette hiérarchisation, on conçoit bien la différence entre les deux définitions : dans le premier cas (déf. 1.1), la chaîne de contrôle passe principalement par l'opérateur ; dans le second cas (déf. 1.2), le contrôle s'effectue principalement via l'ordinateur.

La représentation graphique des définitions données par Sheridan [102] permet d'identifier les différents acteurs du contrôle supervisé (figure 1.1) :

- le système autonome ou semi-autonome (ordinateur, capteur, effecteur),
- l'opérateur humain,
- l'interface Homme-machine (écran, contrôleur).

Cependant, il est à noter que, malgré le découpage des différents modes de contrôle proposé par Sheridan, la classification d'un système n'est pas toujours évidente. Par exemple, dans le cadre du contrôle de drone, l'opérateur peut naviguer entre plusieurs modes de contrôle. En effet, l'opérateur peut assurer la supervision du drone uniquement en définissant un ensemble de waypoints (ensemble de coordonnées géographiques séquencées par lesquelles le drone doit passer) : il ne fait dans ce cas que donner des instructions à l'unité en charge du vol. Mais il peut aussi prendre en charge les commandes de vol du drone, agissant alors directement sur la réalisation de la tâche de pilotage. Nous sommes en présence d'un système, le drone, dont le mode de contrôle peut être le contrôle supervisé ou le contrôle manuel selon que la tâche réalisée est le pilotage du drone ou le suivi d'un plan de vol. Il est intéressant de noter que ces deux tâches peuvent être vues comme deux sous-tâches d'un même processus de traitement de l'information. En effet, si l'on se réfère aux travaux de Parasuraman et Sheridan [87], ce processus peut être séquencé

en quatre étapes : l'acquisition de l'information, le traitement de l'information, la prise de décision, et la mise en œuvre de celle-ci (fig.1.2). Dans notre exemple, la tâche de pilotage est associée à la mise en œuvre de la décision, et la définition de waypoints est associée à la prise de décision. Nous avons donc une classification du mode de contrôle du drone qui est dépendante de la tâche réalisée.

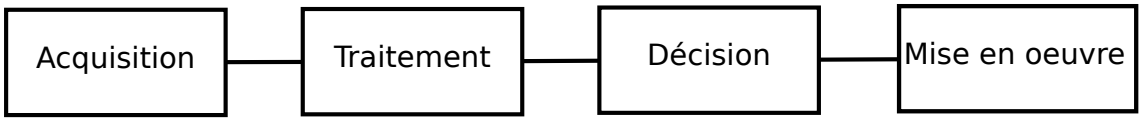

FIG. 1.2 : Processus humain du traitement de l'information en quatre étapes [87].

1.2 MODÈLES DE CONTRÔLE SUPERVISÉ

1.2.1 Contrôle supervisé mono-tâche

Le contrôle supervisé se divise en cinq fonctions :

- La planification (*Plan*) : il s'agit de définir les objectifs et les solutions permettant de les atteindre ;
- La programmation (*Teach*) : l'opérateur doit traduire les objectifs et les stratégies (solutions) en termes d'instructions permettant au système de les réaliser automatiquement ;
- La surveillance des contrôles automatiques (*Monitor*) : l'opérateur surveille l'état du système afin de contrôler le bon déroulement des processus automatiques ;
- L'intervention (*Intervene*) : si le système a fini sa tâche ou que son comportement diverge, l'opérateur doit modifier les instructions initiales afin d'attribuer une nouvelle tâche au système ou corriger son comportement pour revenir à un état opérationnel correct ;
- L'apprentissage (*Learn*) : les données accumulées pendant les procédures permettent par leur analyse d'éviter certaines anomalies à l'avenir. Ces résultats seront exploités dans le cadre des fonctions énumérées précédemment.

Il est intéressant de voir que ces fonctions peuvent être mises en parallèle des travaux de Hess et McNally [50]. Ces derniers décrivent les processus de contrôle à l'aide d'un modèle qui repose sur un ensemble de rebouclages (illustré pour le contrôle de véhicule en figure 1.3). Aux fonctions du contrôle supervisé que sont la planification, la programmation et la surveillance, on peut associer respectivement le guidage, le contrôle et le contrôle du processus issu de l'exemple de Hess et McNally. Les rebouclages liées aux fonctions d'intervention et d'apprentissage du contrôle supervisé sont respectivement la correction des déplacements et la correction de la direction. Suivant cet exemple on peut alors représenter les fonctions du contrôle

supervisé au sein d'un modèle qui repose sur un ensemble de rebouclages (illustré en figure 1.4). Il faut préciser que ce dernier est mono-tâche.

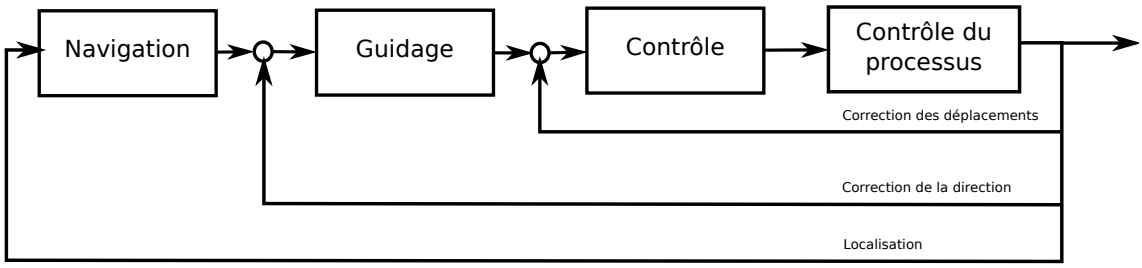

FIG. 1.3 : Navigation, guidage, et contrôle comme un ensemble de rebouclages appliqués au contrôle d'un véhicule.

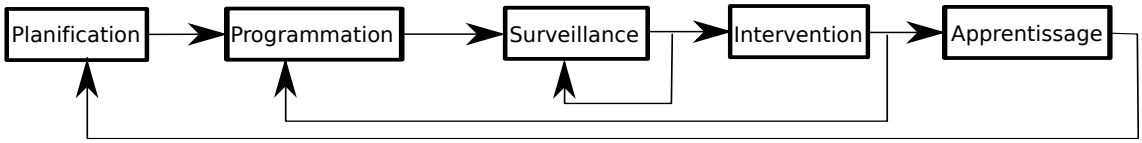

FIG. 1.4 : Modèle basé sur les cinq fonctions du contrôle supervisé (repris de [103]).

Plusieurs éléments sont à relever :

- Tout d'abord, le modèle est mono-tâche. Or, que ce soit pour une centrale nucléaire ou pour le contrôle multi-drones, les tâches à réaliser sont multiples. Prenons l'exemple de la future centrale nucléaire de Flamanville [1]. Le personnel d'exploitation est composé d'un superviseur et de plusieurs opérateurs. Ces derniers ont pour tâches la réalisation des actions manuelles, le contrôle de sûreté, le lancement d'actions correctives et la relation à des opérateurs locaux externes à la salle de contrôle entre autres. Mais il faudrait un modèle de contrôle supervisé qui, en plus de tenir compte de la multiplicité des tâches des opérateurs, tienne compte aussi des capacités de l'opérateur, comme l'illustrent Crandall et Cummings [30] en montrant, entre autres, une corrélation entre l'efficacité de l'opérateur et la fréquence à laquelle il bascule son attention d'une tâche à une autre.
- Par ailleurs, cette corrélation entre l'attention de l'opérateur et son efficacité démontre aussi le besoin d'une prise en compte plus poussée de l'opérateur au sein des modèles de supervision. En effet, toutes les boucles, au sein de ce modèle, ont pour but d'ajuster le comportement du système en fonction de mesures faites sur le processus et sur l'environnement. Elles ne tiennent compte ni d'une erreur éventuelle de la part de l'opérateur, ni de l'opérateur lui-même (comportement vis-à-vis du système, capacités cognitives, allocation d'attention). Il faut donc, si ce n'est au sein du modèle du contrôle supervisé,

que le système possède, a minima, un modèle de l'opérateur pour ajuster son comportement en conséquence.

1.2.2 Contrôle supervisé multi-tâches

Nous avons vu précédemment un modèle mono-tâche du contrôle supervisé. Mais dans le contrôle multi-drones par exemple, chaque drone remplit une mission différente (interception, surveillance, *etc.*). Il faut donc tenir compte dans nos modèles de la pluralité des tâches que supervise un opérateur. Nous allons donc aborder les modèles multi-tâches.

Dans ce but, Sheridan [103] propose un modèle initial illustré par la figure 1.5. Un opérateur peut contrôler plusieurs tâches par l'intermédiaire d'une seule interface Homme-machine (SIH - Système d'interaction côté Homme[1]). Et chaque tâche a sa propre interface logicielle avec le système (SIT - Système d'interaction côté tâche[2]).

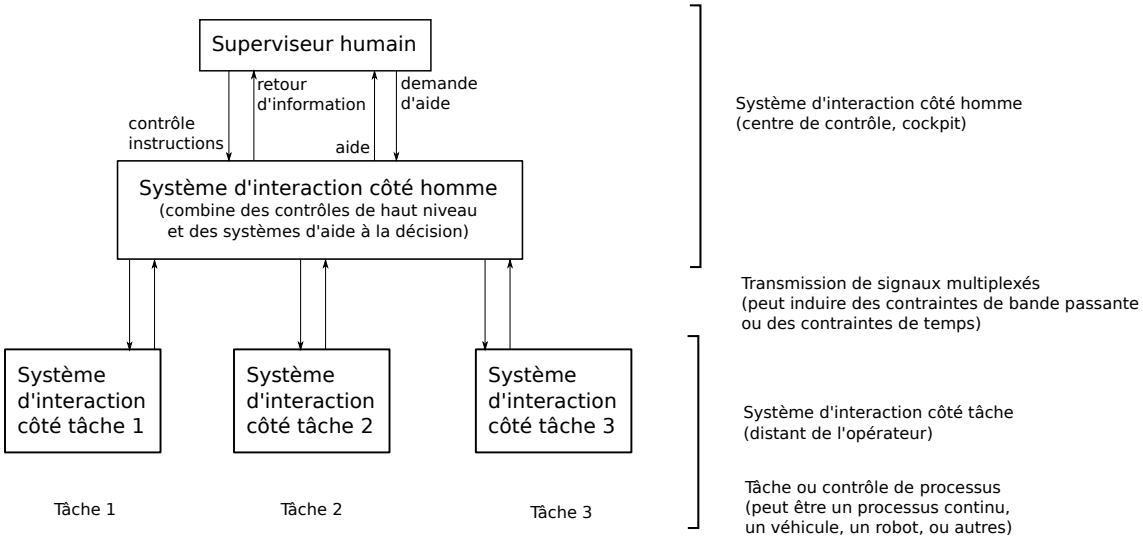

FIG. 1.5 : Cas d'une supervision multi-tâches - repris de [103].

Un modèle de contrôle multi-tâches reposant sur la parallélisation du modèle mono-tâche a également été proposé par Cummings et al. [32]. Ainsi, le modèle de contrôle supervisé mono-tâche a été adapté au contrôle de drone (figure 1.6). En considérant le rebouclage du suivi de mission et de gestion de la charge utile comme commun à toutes les tâches, illustré figure 1.7, on obtient un modèle de contrôle supervisé multi-tâches spécifique au contrôle multi-UVs.

Ces modèles, si nous les comparons au précédent, n'ajoutent que la prise en compte du multi-tâches. Et l'opérateur n'y est toujours pas considéré. Johannsen

[1]HIS, *Human interactive system*
[2]TIC, *Task interactive computer*

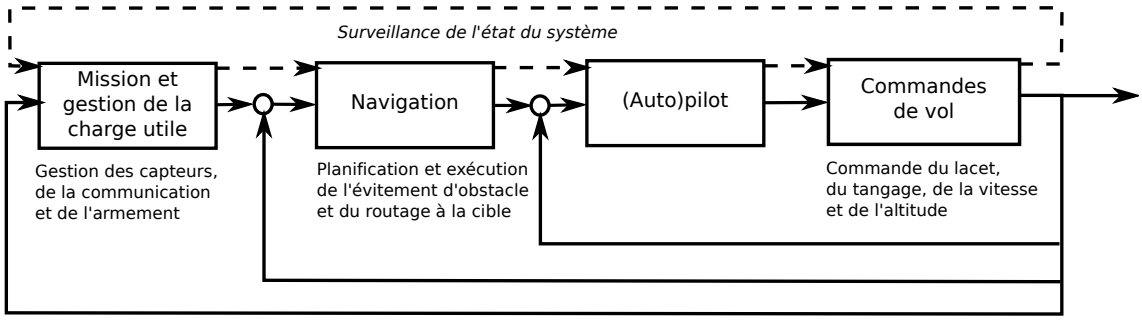

FIG. 1.6 : Navigation, guidage et contrôle comme application du contrôle supervisé à un avion repris de [32].

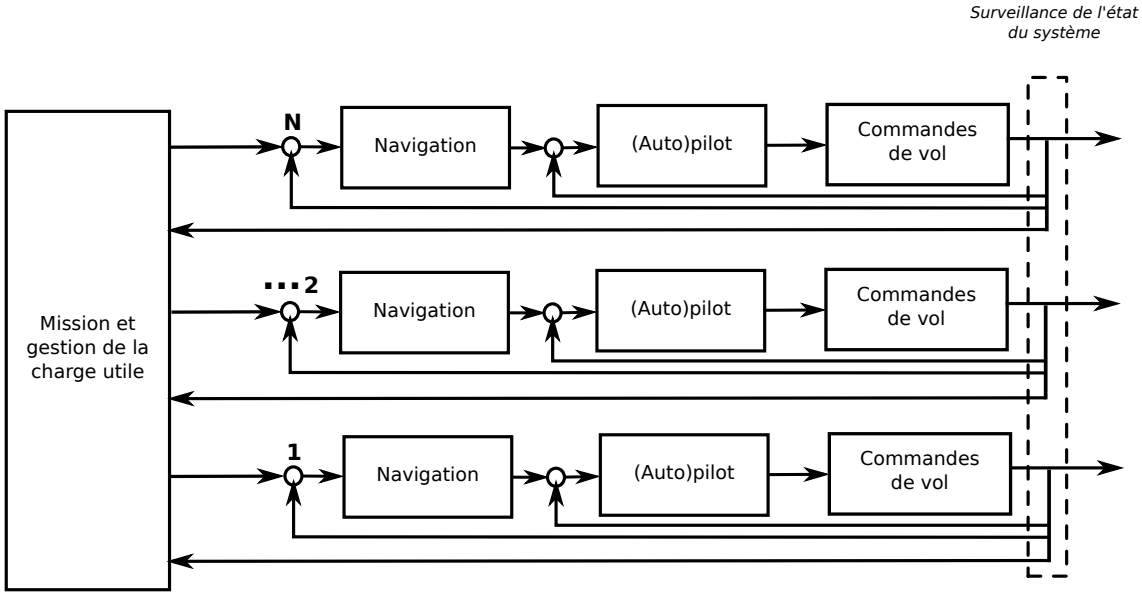

FIG. 1.7 : Contrôle hiérarchique pour de multiple UVs - repris de [32].

répond en partie à ce défaut en modifiant le SIH [58] du modèle de Sheridan. Pour cela, il utilise le paradigme de Rasmussen [92] qui représente le savoir humain selon trois niveaux :

- Le premier niveau (Skill-based) correspond aux habiletés, c'est-à-dire les compétences élémentaires (réflexe, action simple). Il associe à un stimulus une réponse immédiate. En d'autres mots, il ne nécessite pas, ou très peu, de ressource cognitive, pour être exécuté. Par exemple, pédaler sur un vélo relève de ce niveau. En effet, après apprentissage, il n'est pas nécessaire de réfléchir pour réaliser cette tâche.
- Le deuxième niveau (Rule-based) correspond aux règles. Il se caractérise par la mise en œuvre de schémas (actions paramétrées). Un ensemble de stimuli va en-

gendrer l'application d'une règle. Par exemple, dans un établissement scolaire, lors d'un incident, le personnel de l'établissement choisit la règle d'évacuation adaptée à la situation et l'applique. Contrairement au premier niveau la décision n'est pas immédiate et nécessite de sélectionner une règle parmi plusieurs connues (selon le contexte).

– Le troisième niveau (Knowledge-based) correspond aux connaissances. Afin de générer une réponse, l'individu doit faire appel à l'ensemble de ses connaissances pour analyser la situation (raisonnement logique) et définir une réponse adéquate à déclencher. Ce niveau d'action implique une charge mentale bien plus conséquente que les deux premiers niveaux.

Le rapport entre chacun des niveaux et le processus du stimulus à la réaction sont illustrés figure 1.8.

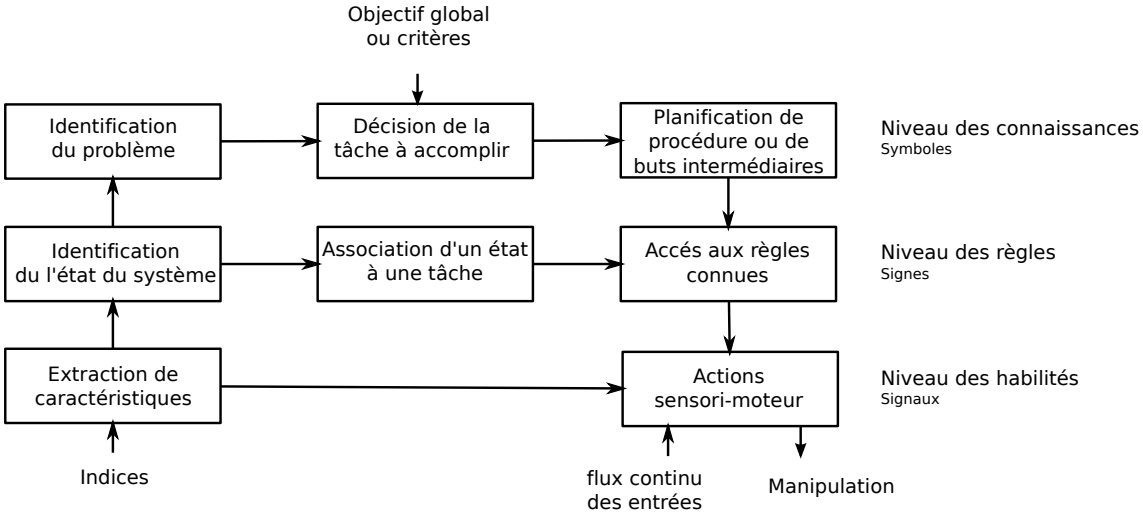

FIG. 1.8 : Paradigme de Rasmussen - repris de [102].

En extension à la couche SIH, Johannsen intègre au système une connaissance sur le mode de fonctionnement de l'opérateur. Cette extension a pour but de modéliser, dans le cadre du paradigme de Rasmussen, le comportement de l'opérateur : stratégie cognitive, traitement de l'information, spécificité des types d'opérateur (technicien, ingénieur, etc.) [56]. Symétriquement, un modèle de l'application (connaissances sur les contrôles bas niveau, modèles des processus, critères d'erreurs) est ajouté en extension de la couche SIT.

Ces connaissances sur l'opérateur et l'application sont utilisées par une unité de gestion de dialogue, permettant la communication entre l'opérateur et le système à l'aide de représentations de type graphique [57]. Les informations sont présentées sous une forme qui est donc plus lisible que de simples séries numériques. Ces modèles sont aussi employés par un système d'aide à la décision qui peut communiquer via cette même boîte de dialogue [59] afin d'aider l'opérateur.

Enfin, Johannsen justifie ses travaux sur les modèles de performances humaines dans le cadre du contrôle supervisé pour la sécurité et la qualité du contrôle, il considère qu'il est plus important d'améliorer l'acceptation des systèmes par leurs utilisateurs, ainsi que la représentation mentale qu'ils s'en font, que de vouloir améliorer la précision des modèles dynamiques [58].

Dans leur ensemble, ces modèles (hormis celui de Johannsen) possèdent deux défauts :

- Premièrement, l'opérateur n'est pas pris en compte. En effet, les modèles introduits précédemment mettent en œuvre des systèmes qui ne tiennent pas compte de l'opérateur, bien que ce dernier fasse partie de la vision d'ensemble.
- Ensuite, la part d'activité de l'opérateur et celle du système ne sont jamais clairement définies. Les différents asservissements de ces modèles décrivent à quel niveau un opérateur peut intervenir pour contrôler et corriger la tâche en cours. Or ils n'abordent pas la répartition d'activité entre l'opérateur et le système. En effet, l'opérateur peut intervenir à plusieurs niveaux de compétence[3] mais ces modèles ne spécifient pas auquel l'opérateur doit intervenir. De plus, lors d'activité multi-tâche, la répartition de l'activité de l'opérateur entre les tâches n'est pas non plus abordée. Si nous reprenons le modèle de Cummings, le contrôle de plusieurs drones est mis en parallèle, mais le modèle ne traite pas de l'activité humaine. Où l'opérateur intervient-il ? Comment distribue-t-il son attention entre les tâches ? Ainsi la répartition des tâches et de l'autorité entre l'opérateur et l'automate doit être clairement étudiée. Il faut donc définir au préalable les capacités d'autonomie des systèmes afin d'aborder les activités du système et celles de l'opérateur ainsi que leurs responsabilités respectives.

1.3 Degré d'autonomie

Sheridan et Verplank [101] distinguent traditionnellement huit niveaux d'autonomie, étendus à dix par Parasuraman [87] illustrés dans le tableau 1.1. Ces dix niveaux vont de la commande manuelle à l'automatisation complète. On y distingue deux sous-ensembles : homme plus ou moins assisté par une machine, et la machine plus ou moins assistée par l'homme, ce qui peut être associé à la définition stricte du contrôle supervisé (cf. déf.1.2).

Toutefois, cette représentation de l'autonomie est incomplète et ne permet pas de décrire toutes les situations. Supposons par exemple un système d'aide à la décision qui proposerait un ensemble de possibilités hiérarchisées et dont la première serait automatiquement sélectionnée après un laps de temps si aucun opérateur n'intervient. L'ensemble de possibilités correspondrait à une autonomie de niveau 2, mais l'auto-sélection relève du niveau 6. Ce type d'automatisme est donc inclassable dans cette taxonomie. C'est pourquoi d'autres taxonomies avec des approches différentes ont été proposées.

[3] Ces niveaux faisant référence au paradigme de Rasmussen vu précédemment

Niveau	Coopération Homme-Machine
10 (Autonomie complète)	L'ordinateur décide de tout, agit seul, ignore l'humain
9	Informe l'humain s'il (l'ordinateur) le décide
8	Informe l'humain s'il (l'humain) le demande
7	Agit puis informe l'humain
6	Accorde un délai pour un éventuel veto avant l'exécution
5	Exécute sa suggestion si l'humain approuve
4	Suggère une possibilité
3	Présente un ensemble restreint de possibilités choisies par la machine
2	Présente un ensemble complet de possibilités
1 (Contrôle manuel)	N'offre aucune assistance

TAB. 1.1 : Niveaux d'automatisations - issu de [87, 101].

Bruni et al. ont proposé, par exemple, *Human-Automation Collaboration Taxonomy* (HACT) [12]. HACT est basé sur un modèle de traitement de l'information collaboratif d'où sont extraits trois acteurs (figure 1.9) : le modérateur qui contrôle le processus de décision, le générateur qui élabore des solutions et le décideur qui sélectionne la solution à mettre en œuvre Une première classification a lieu sur ces rôles. En effet, chacun d'entre eux peut prendre plusieurs niveaux d'autonomie différents : du rôle assumé par l'humain au rôle assumé par l'automate en passant par trois niveaux intermédiaires où le rôle est plus ou moins partagé. Enfin à cette première classification s'ajoutent trois paramètres supplémentaires qui sont :

- la transparence fonctionnelle : la vue que possède l'opérateur sur le fonctionnement de l'automate peut être opaque, partielle ou bien transparente ;
- la transparence de l'information : les informations circulent sous forme de données brutes ou bien sous forme d'agrégation (données traitées et présentées sous forme de graphismes par exemple), ou encore sous forme de mélange de données brutes et agrégées ;
- l'interactivité : cette dernière caractéristique se répartit en deux classes, à savoir le dialogue impliquant une négociation des agents, ou bien la commande sollicitant les actions de façon unilatérale (un agent transmet un ordre, le récepteur peut répondre par une confirmation ou un retour d'information sur le résultat de la commande).

Ainsi toute combinaison d'un modérateur, d'un générateur, d'un décideur et de ces trois caractéristiques nous donne une catégorie spécifique de l'autonomie au sein de la taxonomie de HACT.

Ces deux représentations des degrés d'autonomie sont très différentes et montrent ainsi l'aspect très interprétatif d'une classification de l'autonomie des systèmes, ou

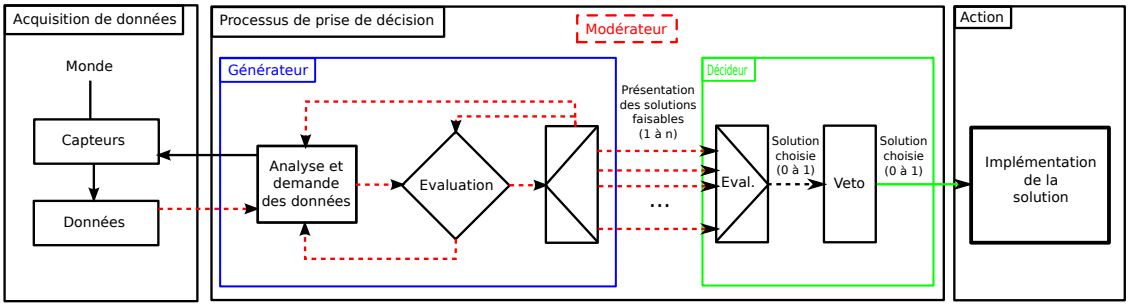

FIG. 1.9 : Modèle collaboratif du traitement de l'information - repris de [12].

de leur mode de contrôle.

1.4 INITIATIVE MIXTE ET PARTAGE D'AUTORITÉ

1.4.1 Définition

Carbonell définit, en 1970, l'initiative mixte (Mixed-initiative) de la façon suivante :

> *A mixed-initiative system is one in which both humans and machines can make contributions to a problem solution, often without being asked explicitly.*

Les actions peuvent être initiées par l'opérateur ou bien la machine. Se présente donc, en fonction des situations, un niveau d'autonomie variable : ce qui signifie que dans la taxonomie des niveaux d'autonomie de Sheridan et al. (figure 1.1), le niveau associé n'est pas toujours le même. Il peut être au niveau 5 à un moment, puis passer au niveau 8 à l'instant d'après. Une initiative de la machine sera par exemple entre les niveaux 6 et 10.

1.4.2 Modèles d'initiative mixte et de partage d'autorité

La notion d'autonomie variable et de partage d'autorité est abordée selon différentes approches. Ainsi nous trouvons des modèles basés sur l'interaction avec une autorité rediscutée en permanence, ou des modèles qui mettent l'autorité de l'homme ou de la machine en avant, l'autre n'intervenant qu'en tant qu'aide. Enfin, nous verrons comment l'initiative mixte peut être abordée sans utiliser la notion de degré d'autonomie.

1.4.2.1 Partage d'autorité et interactions

Une vision du contrôle supervisé mettant en œuvre une autonomie variable basée sur le partage d'autorité a été schématisée par Saget et *al.* [97] (figure 1.10). L'idée

est de modéliser le partage d'autorité avec une représentation d'un opérateur, du système et de la décision finale.

Nous pouvons remarquer qu'un point fort de ce modèle est la considération de représentations réciproques du système et de l'opérateur (absente jusqu'à présent dans les modèles antérieurs). En effet chacun possède une représentation de soi-même mais aussi de l'autre. Ceci permet à chacun de proposer un mode opératoire désiré. Le partage d'autorité a lieu au cours de la sélection finale du mode à partir des propositions. Pour faciliter ce choix final une interaction plus ou moins élaborée entre système et opérateur est nécessaire, afin de faire converger leurs propositions. Une bonne interaction est alors nécessaire. En effet, elle permet d'établir des représentations mutuelles correctes d'une part, mais aussi une inter-compréhension qui facilite la convergence des propositions de chacun.

Enfin, l'utilisation du modèle OODA permet une analyse plus en profondeur de chaque tâche. Ce modèle définit par Boyd [11] permet de décrire le processus nécessaire pour s'adapter à l'évolution constante de notre environnement. Il décrit au sein d'une boucle ouverte les actions qu'un individu doit entreprendre pour s'adapter continuellement. Ainsi ce processus, très similaire à celui du traitement de l'information défini par Parasuraman et Sheridan [87], se décompose en quatre étapes : l'observation, l'orientation, la décision et l'action. Le degré d'autonomie du système est alors défini pour chacune de ces étapes, ce qui permet d'optimiser la réalisation de la tâche dans son ensemble.

1.4.2.2 Initiative mixte : la machine support de l'homme

Rauschert, Meitinger et Schulte [93] ont réalisé une mise en œuvre de l'initiative mixte, dans le domaine des UAVs, à l'aide d'agents cognitifs. Ils implémentent, côté automate, un agent cognitif afin que l'opérateur puisse lui transmettre des ordres de haut niveau[4]. Ceci permet un alignement des modèles de représentation qui réduit la charge cognitive de l'opérateur. De plus, un agent d'aide à la décision est aussi adjoint coté opérateur afin, d'une part, d'attirer l'attention de l'opérateur sur les tâches urgentes et, d'autre part, de prendre l'initiative sur certaines décisions en cas de surcharge cognitive de l'opérateur. Ceci implique donc, de la part de l'agent, la constitution d'un modèle de l'état courant de l'opérateur. À nouveau peuvent être reprochés à cet exemple l'éventualité d'une dépendance de l'opérateur à cette assistance, ainsi qu'un excès de confiance envers celle-ci pouvant provoquer des conséquences négatives d'un point de vue matériel, mais aussi du point de vue de la tâche (échec partiel ou total).

[4]Correspondant à la manipulation de symbole si l'on se réfère au paradigme de Rasmussen.

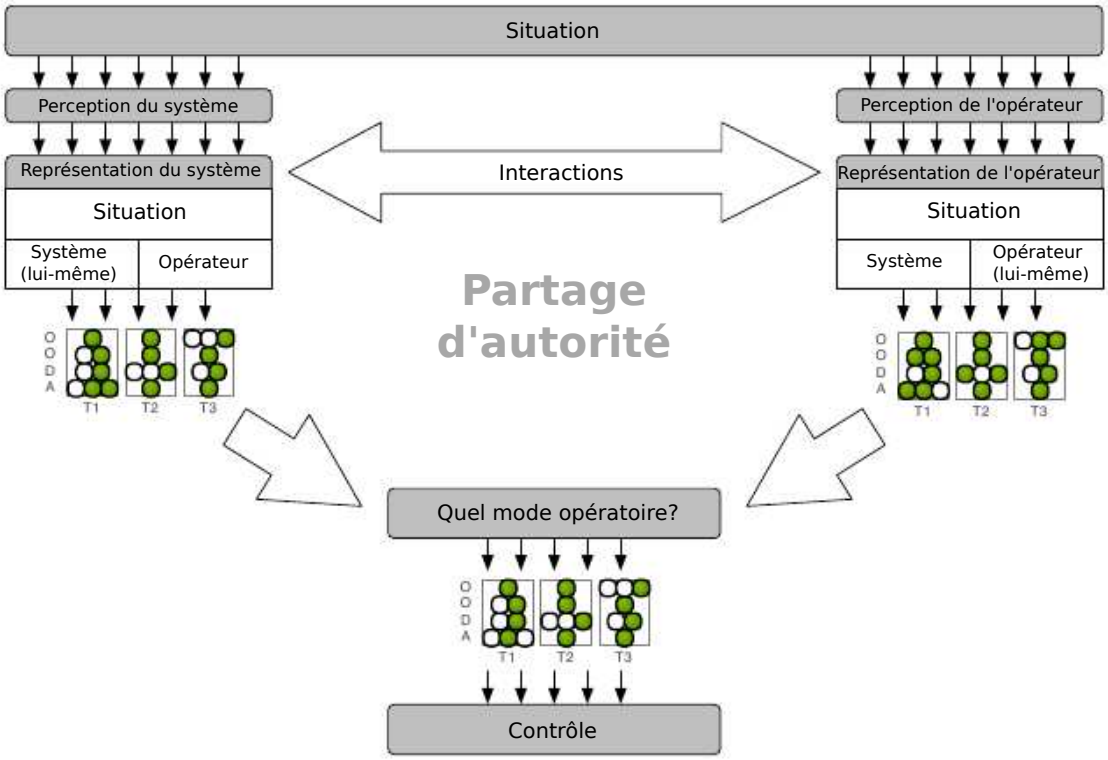

FIG. 1.10 : Partage d'autorité selon Saget et *al.* - issu de [97].

1.4.2.3 Iniative mixte : l'homme support de la machine

Un certain nombre d'applications de l'autonomie ajustable – autonomie variable et ajustée par l'opérateur – ont pour objectif l'augmentation du nombre d'engins autonomes que peut contrôler un seul et unique opérateur. Ainsi un modèle réalisé par Mouaddib [79] part du principe que tous les engins sont parfaitement autonomes, et l'opérateur n'intervient que sur requête de ces derniers pour les aider. Ce modèle a l'avantage de permettre le contrôle d'un nombre important de robots mais ne tient plus compte de l'opérateur. Celui-ci n'a, en effet, qu'une faible conscience de la situation concernant le robot ayant requis son intervention.

1.4.2.4 Partage d'autorité basé sur la notion de ressources

Au lieu d'aborder le partage d'autorité par la définition d'une autonomie variable, Mercier *et al.* propose un modèle basé sur la notion de ressource [74]. Cette approche cherche à résoudre les conflits décisionnels entre les opérateurs et les systèmes autonomes. En effet, il peut arriver que les algorithmes d'un véhicule décident de faire tourner ce dernier à gauche tandis que l'opérateur intervient pour le faire tourner à droite. Afin de résoudre ce conflit, Mercier *et al.* considère le système de

direction comme une ressource partagée à laquelle l'opérateur et les algorithmes ne peuvent accéder que de façon exclusive.

Les ressources sont modélisées à l'aide de réseaux de Petri qui décrivent à la fois son état (utilisé ou disponible) et ses propriétés (partageable, préemptable). Les conflits sont détectés par l'observation d'inconsistances au sein de ces réseaux, par exemple, une ressource non partageable allouée à deux utilisateurs. Des procédures sont alors mises en œuvre afin de les résoudre.

Ce modèle est basé essentiellement sur un ensemble de règles qui gèrent l'accès aux ressources. Or nous avons, jusqu'à présent, insisté un certain nombre de fois sur l'importance d'une prise en compte de l'opérateur. Dans ce dernier modèle, celui-ci est réduit au rôle de simple agent en concurrence avec le système pour accéder aux ressources.

Nous avons pu voir à travers un certain nombre d'exemples récents que la mise en œuvre de solutions à autonomie variable pour le contrôle supervisé permet de tenir davantage compte de l'opérateur. En effet, ces modèles permettent de décharger l'opérateur d'une partie de ses tâches lorsque celui-ci est en surcharge. Ceci se produit à l'initiative du système (Rauschert), par le dialogue (Saget). Or tous les incidents ne sont pas nécessairement liés à une surcharge de l'opérateur. En effet, bien que soient prises en compte les capacités de l'opérateur, un autre facteur est délaissé : l'attention de l'opérateur. Par exemple, lors du contrôle multi-drones, l'opérateur partage son attention entre les différents UVs. Pour le contrôle de l'un des drones, des erreurs peuvent intervenir non pas parce que l'opérateur avait une trop forte charge cognitive mais tout simplement parce que son attention n'était pas portée au bon endroit. Les modèles du contrôle supervisé ne doivent pas tenir uniquement compte des capacités de l'opérateur mais de tout un ensemble de problématiques liées à l'automatisation et à son usage. Ces problèmes concernent d'une part les capacités des opérateurs mais aussi leur conscience de la situation, les biais de décision entre autres.

1.5 Problèmes sous-jacents au contrôle supervisé

Les critiques soulevées précédemment se concrétisent dans des réflexions de Parasuraman [85, 88] sur l'automatisation et son usage. Nous commencerons par observer les problèmes de charge cognitive, de conscience de la situation et de dégradation de compétence pour ensuite aborder leurs conséquences dans l'usage des automates.

1.5.1 Problèmes liés à l'automatisation

L'automatisation des tâches apporte un certain nombre de problèmes, et ce, quel que soit son degré. C'est pourquoi nous allons aborder dans cette section l'impact

négatif qu'a l'automatisation sur la charge cognitive, sur la conscience de la situation ainsi que sur les compétences d'un opérateur.

1.5.1.1 Charge cognitive

La charge cognitive peut être décrite comme la fonction de relation entre les ressources mentales nécessaires pour une tâche et les ressources disponibles de l'opérateur humain, selon Parasuraman et *al.* [89]. Accroître le niveau d'automatisation des systèmes de façon efficace permettrait, selon Kantowitz [61], de réduire la charge cognitive des pilotes pour l'aviation. Cependant, la simple augmentation du degré d'automatisation n'améliore pas nécessairement les performances du système, en termes de réduction de charge de cognitive. Dans des situations anormales et inattendues, l'automatisation peut échouer ; pouvant atteindre des proportions catastrophiques tandis que l'opérateur ne peut pas prendre en charge la tâche concernée (Cummings et *al.* [34]). Le manque de transparence et de confiance envers le système peut potentiellement mener à l'augmentation de la charge cognitive pour l'opérateur, en voulant déterminer si l'automate fonctionne correctement ou non et si une intervention est requise. Raeth et Reising [91] ont étudié l'évolution de la confiance et de la charge cognitive d'un pilote de combat lors d'une mission d'attaque. Ils montrent ainsi que la confiance est de plus en plus nécessaire au fur et à mesure que la charge cognitive de l'opérateur augmente. En effet, lors de l'attaque, le pilote n'a plus le temps de réfléchir aux informations transmises par le système. Il doit faire confiance, et donc déléguer ses tâches le plus possible à l'automate. En revanche, sur les phases de transit, le pilote aura ses propres ressources disponibles pour contrôler les opérations de vol et n'aura plus à se reposer fortement sur les systèmes automatiques. Il est donc important, dans ces phases plus calmes, de construire le lien de confiance.

Certains systèmes automatiques ont pour but de formater les informations dans l'optique de faciliter leur intégration par l'opérateur. Nous pouvons citer comme exemple le système de prédictions de trajectoires en aéronautique réalisé par Morphew et Wickens [76] qui a pour but de permettre à un pilote d'anticiper les évolutions du trafic aérien environnant. Plus ce système apporte des informations prédictives sur l'évolution du trafic et plus la charge cognitive du pilote diminue. Le système de fusion d'informations réalisé par O'Hara [84] dans le cadre du désencombrement informatif d'une interface de contrôle est un autre exemple. L'objectif est de diminuer la quantité d'information visuelle affichée. Ainsi au lieu d'indiquer sur une carte l'emplacement de chaque navire de pêche d'une flottille, l'interface va regrouper ces informations et créer une information indiquant la zone de localisation de la flottille. Nous nous intéressons donc principalement à l'impact de la transparence de l'information (paramètre de la taxonomie HACT vu en 1.3 sur la charge cognitive de l'opérateur). Mais comme le souligne Legras [67] l'effet peut être négatif : en effet, si le modèle de représentation des données s'éloigne de celui de l'opérateur, les conséquences seront une augmentation de sa charge cognitive au lieu d'une diminution. Il

faut bien sûr avoir à l'esprit que les capacités humaines dans le traitement de l'information sont limitées (nombre limité d'éléments informatifs simultanés : capacité de la mémoire à court terme), d'autant plus lorsqu'elles sont sollicitées dans un cadre multi-tâches (démultiplication des paramètres). Ce problème ne se restreint pas à la forme de l'information mais s'étend aussi à son volume : un système sous-automatisé où la quantité de tâches dévouées à l'opérateur est trop importante peut engendrer le même type de surcharge cognitive.

1.5.1.2 Perte de conscience de la situation

La conscience de la situation[5] est "la perception des éléments environnants dans un espace-temps donné, la compréhension de leur sens et la projection de leur état futur dans un avenir proche"[6] (Endsley et Garland [45]). La conscience de la situation n'est pas un choix ou un acte basé sur les informations présentes, ou une conséquence d'un diagnostic, ni la mesure des performances d'un opérateur humain, mais plutôt le modèle mental que possède l'opérateur de la situation courante et à venir. Des études montrent que, dans les opérations d'UAV, la conscience de la situation varie inversement au niveau d'automatisation. Drury et Scott développent le diagramme de la figure 1.11 pour illustrer ce concept [42].

Ce diagramme montre que, lorsque le niveau d'automatisation d'un UAV augmente, le niveau de détail d'informations et donc la conscience de la situation associée de l'opérateur au regard de l'UAV tendent à décroître. Par exemple, l'UAV Global Hawk est un véhicule aérien hautement automatisé qui peut voler entre deux points de contrôle pré-programmés, et, par conséquence, l'opérateur humain a besoin d'une conscience de la situation des informations relatives au vol de l'engin moindre que les opérateurs de Predator poursuivant une cible mobile.

Des recherches antérieures ont décomposé la conscience de la situation en trois niveaux (Endsley et Garland [45]) :

- Niveau 1 - perception de l'information,
- Niveau 2 - compréhension de la situation courante,
- Niveau 3 - projection des états futurs.

Parasuraman et *al.* soulignent le fait que ces niveaux peuvent être précisément mesurés pour fournir une compréhension de la conscience de la situation qu'un opérateur aura en utilisant le système. Par exemple, les outils de scan visuel sont commercialement disponibles pour aider à l'évaluation de niveau 1 de la conscience de la situation [89].

Un opérateur peut donc perdre conscience de la situation par manque d'informations [67], ces dernières étant gérées automatiquement par le système. Il peut ainsi en ressortir une mauvaise représentation de la situation. C'est pourquoi il est important soit de remonter un plus grand nombre de données sans que cela n'influence

[5] Situation awareness

[6] the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future

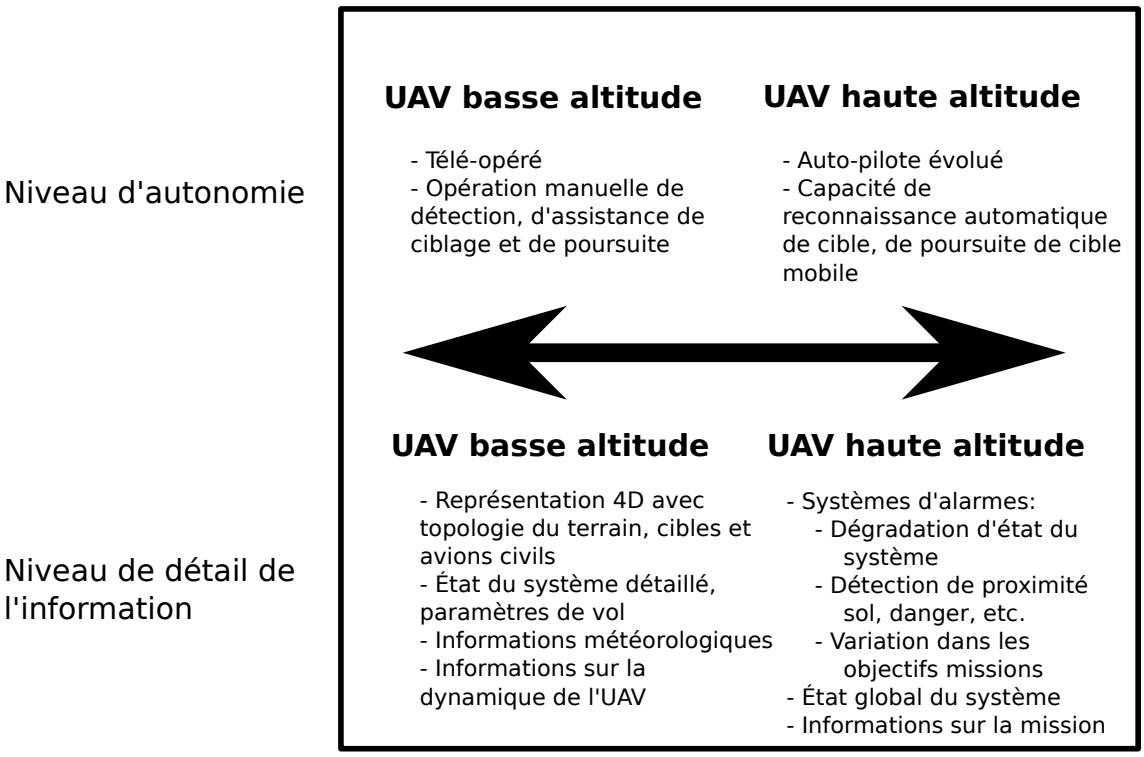

FIG. 1.11 : Conscience de la situation selon le degré d'automatisation d'un UAV - issue de [42].

trop la charge cognitive de l'opérateur, soit de diminuer l'autonomie du système afin d'obtenir une participation active de l'opérateur, l'incitant ainsi à avoir un suivi de situation plus actualisé.

1.5.1.3 Dégradation de compétences

Tout comme le suivi de situation, un automatisme trop important peut enclencher une dégradation des compétences de l'opérateur. En effet, il a été démontré par Kaber [60] qu'une défaillance sur un automate est d'autant plus grave que ce dernier est autonome. La non participation de l'opérateur entraîne un oubli de son savoir-faire. Il faut donc le faire participer à la tâche (en respect de sa charge cognitive).

Nous venons de voir que l'automatisation apporte un certain nombre de problèmes en détachant l'opérateur de ses tâches. En effet, celui-ci déconnecté de ces dernières perd deux choses : la conscience de la situation, et ses compétences. De plus, l'automatisation bien qu'elle permette à un opérateur d'effectuer plus de tâche à la fois, peut, si elle est mal faite, avoir l'effet inverse. Ce dernier point peut aussi

être lié à l'usage de l'automate — qui est conçu correctement mais pas pour l'usage qu'en fait l'opérateur.

1.5.2 Problèmes liés à l'usage de l'automate

Dans les problèmes liés à l'usage d'un automate, deux cas sont à aborder : le mauvais usage (*misuse*) et le non usage (*disuse*).

1.5.2.1 Mauvais usage

Les automates sont habituellement fiables et prédictibles. Cependant ils peuvent aboutir à des échecs ou à un comportement non prévu. De telles occurrences sont généralement peu fréquentes et n'incitent pas les utilisateurs à douter de l'automate. Nous pouvons nous demander alors si la confiance donnée par les utilisateurs n'est pas excessive. En effet, un certain nombre d'accidents du trafic aérien est dû à une erreur de supervision de la part du pilote, qui n'a pas pris en compte un changement d'état du système. Par exemple, un pilote a fait s'écraser son avion faute d'avoir remarqué le désengagement du pilote automatique[7]. En effet, le pilotage trop confiant vis-à-vis du pilote automatique n'a pas surveillé ce dernier, ne pouvant ainsi remarquer son changement d'état et ainsi prendre les décisions nécessaires à temps. De façon plus générale, de telles situations peuvent être liées deux facteurs : les biais de décision et les erreurs de supervision.

Biais de décision :

Dans un cadre incertain, un humain peut déployer une grande variété d'heuristiques pour prendre une décision aboutissant à des biais de décision (sous-estimer l'influence d'un échantillon, en termes statistiques, ou avoir un excès de confiance dans sa décision). Ces heuristiques, étudiées par Tversky et Kahneman [111, 109], sont utilisées quotidiennement, en raison de économie cognitive — et ce, même chez les experts [110]. Bien que les heuristiques soient une alternative aux méthodes analytiques et normatives, elles peuvent mener à des biais dégradant la performance décisionnelle.

Selon Mosier et Skitka [77], les automates qui fournissent une aide à la décision renforcent la tendance humaine à l'usage de telles heuristiques aboutissant à un biais d'automatisme (un mauvais usage de l'automate). Ceci se traduit par des erreurs que l'opérateur n'a pas relevées ou qui n'ont pas été portées à son attention par l'aide décisionnelle. On retrouve ici une critique formulée précédemment sur les travaux de Rauschert et *al.* où une assistance cognitive artificielle est utilisée pour aider un opérateur à sa tâche de supervision [93].

Mosier et Skitka montrent que se fier aux décisions de l'automate diminue l'attention sur les contradictions des sources d'informations. En effet, un opérateur aura

[7]National Transportation Safety Board (1973). *Eastern Airlines L-1011. Miami. Florida. 20 December 1972* (Rep. NTSB-AAR-73-14). Washington, DC.

tendance à prendre les informations issues de l'automate sans se poser de question. Il ne contrôle plus l'information en l'entrecroisant avec d'autres sources, mais il se contentera par exemple de lire l'information "jauge pleine" sans la vérifier à l'aide d'une caméra interne à la jauge. Ceci va donc générer des conflits entre ce que réalise l'automate (ravitailler) et ce qu'attend l'opérateur (continuer la mission), conflits qui seront tout simplement ignorés. Mosier va plus loin dans son étude et précise que même des utilisateurs experts agissent de la même manière. [78].

Cohen et *al.* [24] soulignent aussi l'importance de la relation entre les contraintes temporelles et la confiance dans la prise de décision. Ils proposent un compromis sur la décision en fonction de la confiance (fig.1.12). Si la confiance d'un opérateur dans la conclusion d'un système d'aide tombe dans la région supérieure, alors l'utilisateur pourra simplement accepter la conclusion sans prendre trop de temps. Si la confiance correspond à la région inférieure, l'utilisateur pourra rejeter l'aide sans trop de délai. Enfin, si la confiance tombe dans la région intermédiaire, il est alors intéressant pour l'utilisateur de prendre son temps pour décider de la suite des actions.

L'incertitude et les contraintes temporelles influencent la fiabilité des décisions. Les contraintes temporelles, en déterminant les frontières supérieures et inférieures, agissent comme paramètres principaux du coût d'un délai décisionnel (en termes de conséquences). L'action est plus impérative quand le coût d'un délai est élevé, même avec une grande incertitude concernant la confiance. De plus, quand la pression temporelle augmente, les frontières supérieures et inférieures tendent l'une vers l'autre, indiquant le besoin de prendre une décision avec moins de vérification à un même niveau de confiance (Cohen et *al.* [24]).

Erreurs de supervision. :

Les erreurs humaines de supervision sont une deuxième cause au mauvais usage des automates. Elles constituent également les prémices des biais de décision. En effet les opérateurs peuvent ne pas avoir suffisamment contrôlé les informations concernant l'automate avant de prendre une décision. Le biais n'est plus lié à la prise de la décision mais aux mauvaises informations utilisées pour prendre cette décision.

Il a été montré par Hilburn et *al.* [51] que les opérateurs du contrôle du trafic aérien non-assistés étaient plus prompts à détecter les dysfonctionnements secondaires[8] que des opérateurs assistés par un système d'aide. Le risque d'inattention est d'autant plus important que l'autonomie du système augmente et/ou que le nombre de sous-systèmes augmente (aide à la décision, partage d'autorité, alarmes, etc.). Les conséquences d'une mauvaise surveillance des systèmes peuvent induire un excès de confiance dans la fiabilité de l'automate. Ceci influence alors les décisions de l'opérateur et génère donc des biais de décisions qui aboutissent à une mauvaise utilisation des systèmes.

Par ailleurs, Lee et Moray [65] montrent, à l'aide d'une simulation de contrôle de processus, que si la confiance de l'opérateur envers le système est prépondérante

[8] n'ayant pas d'incidence sur le contrôle du trafic ; par exemple le non-acquittement verbal d'une directive, bien qu'elle soit exécutée, est un dysfonctionnement secondaire

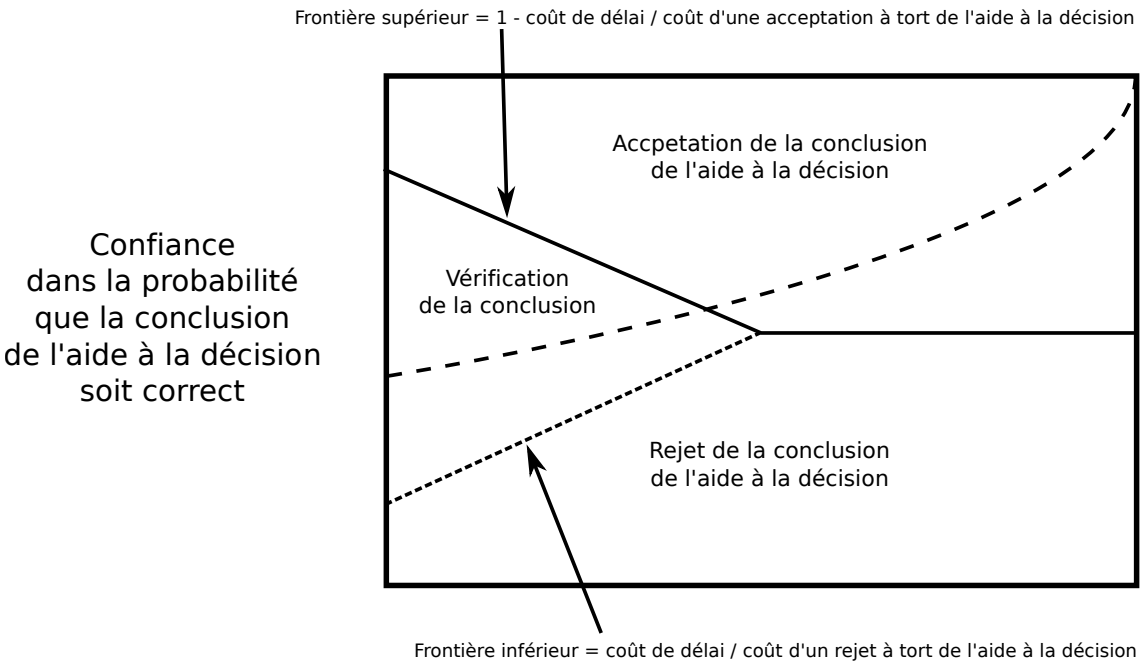

FIG. 1.12 : Acceptation (ou Rejet) d'une aide à la décision selon le niveau de confiance et le temps écoulé - issue de [24]. Plus le temps passe, plus l'aide sera acceptée (ou rejetée) par décision de confiance (ou de méfiance) ne pouvant être vérifiée. La ligne pointillée supérieure représente le niveau typique de confiance d'un opérateur envers un automatisme particulier au cours du temps.

par rapport à la confiance qu'il a en lui-même, celui-ci risque d'abuser de l'usage du système. Cela peut aboutir à un mauvais usage du système, c'est-à-dire à son usage dans une situation inadaptée à sa conception. Inversement, une confiance en soi prépondérante peut conduire à une sous-utilisation du système. Des études conduites par Kantowitz [62] et Riley [95] montre que le degré de familiarité avec le contexte (Kantowitz) ou le système (Riley) influence la perception de la fiabilité du système ainsi que le degré de confiance envers le système. Cela va accentuer la prépondérance de la confiance ou de la confiance en soi, et ainsi augmenter le risque d'un mauvais usage du système.

1.5.2.2 Non usage

Lorsque de nouveaux automatismes sont déployés, leur acceptation par l'opérateur n'est pas systématique. En effet, dans un premier temps, l'opérateur peut manifester de l'appréhension de l'automatisme, voire de la méfiance. Par la suite, au fur et à mesure que l'opérateur acquiert de l'expérience avec le nouveau système, un automate fiable et précis tendra, normalement, à gagner la confiance auprès de l'opérateur. Mais ce n'est pas toujours le cas. Il a été observé avec des systèmes d'alertes

automatisés (système anti-collision par exemple) que certains ne gagnaient pas la confiance de leur opérateur. En effet, à cause de la nécessité de ne pas omettre une seule alarme, le grand nombre de fausses alarmes stoppe tout processus de construction de confiance et d'acceptation. Si le système est imposé à l'opérateur, ces derniers recourent à des contournements plus ou moins créatifs de l'automatisme (Satchell [98]).

Bien que ce processus soit décrit pour les systèmes d'alertes, nous pouvons facilement imaginer sa transposition à un système automatique dans son ensemble. Ainsi l'automatisation d'une nouvelle tâche peut conduire à un rejet de l'automate par l'opérateur, ce dernier considérant qu'il sera plus efficace d'effectuer lui-même la tâche plutôt que de corriger l'automate (si nécessaire). Pour ces raisons, un automatisme laissant un accès à un contrôle direct de la tâche pourra être ignoré par son opérateur.

1.5.3 Contrôle supervisé et confiance

Nous avons vu précédemment un certain nombre de problèmes liés à l'automatisation. Ainsi la charge cognitive bien que diminuée par l'automatisation, ne peut avoir lieu que si l'opérateur fait suffisamment confiance au système pour s'y fier. Mais, même si l'automatisation est un gain pour l'opérateur, l'usage qu'il en fait n'est pas toujours approprié. En effet, les décisions de l'opérateur sont sujettes à des biais. Ces derniers seront d'autant plus importants que l'opérateur subira des contraintes temporelles. A nouveau la confiance joue un rôle dans ce processus en guidant de plus en plus la décision de l'opérateur à mesure que le temps presse. La confiance a généralement un rôle plus ou moins important au sein des différentes problématiques du contrôle supervisé et de l'automatisation.

Parasuraman a montré [86] qu'un système très fiable peut engendrer chez l'opérateur un excès de confiance, l'induisant en erreur sur les capacités réelles du système. Alberdi montre qu'avec des systèmes d'aide à la décision, les décisions des opérateurs peuvent être pires que sans cette aide [3]. Un retour d'informations sur le comportement de ce dernier pourrait améliorer l'image qu'a l'opérateur du système (montré par Wickens [112]), mais ceci entre en conflit avec le problème de charge cognitive. En effet, nous augmentons la quantité d'informations à traiter par l'opérateur, d'où les travaux de Raeth et Reising qui tiennent compte non seulement d'une évaluation de la confiance mais aussi de la charge cognitive [91].

Par ailleurs, l'abus d'utilisation de l'automate, pour des capacités ou performances qu'il n'a pas, biaise les perceptions ainsi que les décisions d'un opérateur (*cf.* 1.5.2.1). Selon Langers l'homme réalise une validation cognitive prématurée[9] qui affecte son attitude envers l'automate [63]. Cet excès de confiance envers un système augmente d'autant plus le risque d'erreurs de supervision que le système est autonome. Ainsi, il existe un bouclage entre perception, décision et confiance,

[9] premature cognitive commitment

dans lequel le niveau de confiance est primordial si l'on veut tendre vers un usage optimum de l'automate. Ainsi les biais de décisions, les erreurs de supervision et autres problèmes du contrôle supervisé seront moins nombreux.

2 Confiance

Nous avons vu que le niveau de confiance de l'opérateur envers l'automate est un élément clé du contrôle supervisé. En effet, selon que l'opérateur a trop ou n'a pas suffisamment confiance, les conséquences peuvent être graves d'un point de vue matériel mais aussi humain. Nous proposons de mesurer la confiance d'une part pour permettre la prédiction de ces situations, et d'autre part pour permettre leurs résolutions par anticipation.

Tout d'abord, il est important de souligner la différence de vocabulaire entre le français et l'anglais, qui possède un plus grand champ lexical autour de la confiance. En effet, dans les références anglo-saxonnes citées dans ce manuscrit, la confiance est traduite par *trust*, *confidence* ou *reliance*. Nous reprenons la traduction de Quéré [70] pour les deux premiers termes, à savoir : confiance décidée (*trust*) et confiance assurée (*confidence*). Et nous proposons des traductions spécifiques pour chacun des termes anglais dans le tableau 2.1.

Anglais	**Français**
trust	confiance décidée, confiance qui est justifiée par une réflexion
confidence	confiance assurée, confiance qui est instinctive
reliance	confiance de fiabilité, l'acte de si fier à quelqu'un, quelque chose
trustor	mandataire, celui qui fait confiance
trustee, trusted	fiduciaire, celui à qui on fait confiance
dependability	fiabilité, la possibilité de déléguer en toute confiance
reliability	fiabilité technique, lié aux capacités et aux compétences
expectancy	espérance, attente, c'est ce que l'on pense obtenir en faisant confiance

TAB. 2.1 : Traduction proposée des termes anglais en français.

L'objet de confiance : est le fiduciaire de la confiance, le récipiendaire. Par exemple, un système tel qu'un UAV est composé de plusieurs fonctionnalités distinctes : il faut donc déterminer si la confiance s'applique au système dans sa globalité ou à une fonctionnalité particulière. Les travaux réalisés sur les mauvais usages et non usages des automates ont toujours porté sur des fonctionnalités précises, telles que la détection anti-collision (pour les avions). Nous considérons donc que la confiance s'applique à chaque fonctionnalité du système. Par exemple un opérateur peut éprouver une méfiance vis-à-vis du système de navigation automatique

d'un drone mais faire pleinement confiance à la détection automatique de cible. La confiance (réciproquement la méfiance[1]) vis-à-vis d'une fonction n'implique nullement qu'un opérateur fasse confiance aux autres fonctionnalités du système. Dans le cadre des systèmes à autonomie variable, il faut même affiner l'objet de confiance avec un mode opératoire particulier, c'est-à-dire à une fonctionnalité configurée à un niveau d'autonomie donnée. Par exemple, les systèmes de navigation peuvent être complètement autonome (autonomie complète) ou suivre un pattern défini par l'opérateur (autonomie partielle, niveau 2 à 9 sur l'échelle de Sheridan). La confiance dans le système de navigation pourra donc être différent selon son degré d'autonomie. Ainsi nous évaluerons la confiance de l'opérateur pour une fonctionnalité dans un mode opératoire donné. Il pourra être intéressant, par la suite, d'étudier l'existence d'une corrélation entre la confiance associée à chaque élément du système et la confiance globale du système.

2.1 Contexte

Les travaux sur la confiance débutent en psychologie avec Deutsch [40]. Ce dernier étudie les circonstances qui amènent à une décision basée sur confiance. Par la suite deux chercheurs développent parallèlement leurs visions de la confiance. Le premier, Barber, s'intéresse principalement aux facteurs de la confiance [6]. Quant au second, Luhmann, il présente la confiance comme un moyen de simplification de notre compréhension du monde afin de faciliter nos décisions.

Ces travaux sont à la base de la plupart des études sur la confiance en l'automatisation que ce soit pour l'identification ou la modélisation des relations des facteurs de la confiance (Muir [80], Lee [66], Bhattacharya [7], Cohen [24]).

La confiance dans ces travaux est un moyen de décision. D'ailleurs certains se sont attachés à décrire le processus de ces décisions dans le cadre des relations Homme-machine (Bisantz [9], Dzindolet[43, 44], Gao [49]), quand d'autres s'en sont inspirés pour modéliser la confiance entre agents dans le cadre des systèmes multi-agents (Marsh [71], Castelfranchi [13, 14, 15], Sutcliffe [107]).

Enfin il est important de noter qu'un lien a été établi par Jian [54] entre la confiance Homme-Homme et la confiance Homme-machine. En effet, selon son étude (développé au paragraphe 2.5.1.1), la confiance vis-à-vis de l'homme ou de la machine est perçu de la même façon. On peut donc considérer que les études sur la confiance interpersonnelle sont transposables à la confiance Homme-machine.

Nous débuterons notre réflection à partir de la mise en parallèle des définitions de la confiance faites par Bhattacharya [7] et Lee [66].

[1] *mistrust*, soit le contraire de la confiance décidée

2.2 Définitions de la confiance

Les définitions de la confiance sont nombreuses et varient selon le domaine de recherche (psychologie, sociologie, économie, etc...) et leurs auteurs. Bhattacharya et *al.* [7] ont établi un certain nombre de critères pour définir la confiance :

1. "La confiance existe dans un environnement incertain et risqué, et est inutile lorsqu'il y a certitude." Bhattacharya parle ici d'une confiance décidée, c'est-à-dire d'une confiance qui repose sur une observation et une analyse de notre environnement pour prendre une décision : *trust*.
2. "La confiance représente une espérance du mandataire qui reflète la prédictibilité." Nous parlons toujours d'une confiance décidée qui s'applique à un mandataire à savoir la personne ou le système à qui on fait confiance : *trustee, trusted*. De plus cette confiance est donnée dans l'espérance d'obtenir un gain en retour : *expectancy*.
3. "Toute définition de la confiance doit prendre en compte l'importance et la force de la confiance, où la force représente le degré de confiance du mandataire dans l'espérance d'un résultat, et où l'importance représente la valorisation de ce résultat."
4. "La confiance est spécifique à une situation, à une personne."
5. "La confiance reflète le degré d'espérance d'un résultat positif."

La confiance a été définie par Lee et See [66] comme : "l'attitude d'un agent qui aide un individu à atteindre ses objectifs dans une situation caractérisée par l'incertitude et la vulnérabilité."[2], l'agent pouvant être un automate ou un humain. En reprenant cette phrase au vu des critères de Bhattacharya, nous constatons que ces définitions sont en partie différentes. Nous retrouvons la notion d'incertitude à la fois dans la définition de Lee "une situation caractérisée par la vulnérabilité", et dans le premier critère de Bhattacharya. Mais les notions d'espérance et de valorisation que l'on retrouve dans les critères 2, 3 et 5 n'apparaissent pas aussi clairement dans la définition de Lee. Nous pouvons penser que la notion de "vulnérabilité" dans la définition de Lee, si elle ne retranscrit pas une force ou un degré d'attente, traduit au minimum une notion d'espérance. En effet, on peut considérer que plus on se sent vulnérable, plus notre espérance est faible. On retrouverait donc le deuxième critère de Bhattacharya mais amputé de la notion de prédictibilité. Enfin la définition est en accord avec le quatrième critère de Bhattacharya. La confiance est spécifique à une personne ("un agent", Lee) et à une situation ("une situation", Lee).

Bien qu'en français les mots soient semblables, une distinction est faite entre *trust* et *reliance*. En effet, selon Lee et See : "la confiance est entre les connaissances qui concernent les caractéristiques de l'automate et l'intention de se fier à l'automate"[3].

[2] the attitude that an agent will help achieve an individual's goal in a situation characterized by uncertainty and vulnerability.

[3] Trust stands between beliefs about the characteristics of the automation and the intent to rely on the automation

De plus, ils ajoutent “la confiance guide - mais ne détermine pas complètement - la fiabilité”[4].

2.2.1 Facteurs de la confiance

La confiance dépend de nombreux facteurs. Telle que définie par Bhattacharya et *al.*, la confiance est spécifique à une situation, un individu et varie au cours du temps. Mais malgré tout, un certain nombre de facteurs restent constants au sein de la littérature scientifique.

2.2.1.1 Environnement et contexte

L’environnement au sein duquel est utilisé l’automate joue un rôle important dans le besoin de l’opérateur à lui faire confiance. En effet, selon Luhmann, la confiance est une façon de réduire la complexité de l’environnement [69]. Ainsi, un opérateur utilisant un automate dans un environnement complexe ressentira un besoin de simplification pour appréhender la situation et par extension un besoin de confiance. Ceci fait d’ailleurs parfaitement écho aux travaux de Tversky et *al.* [111, 109] sur l’usage d’heuristiques dans le raisonnement humain. Selon Bhattacharya et *al.*, la confiance n’est pas nécessaire lorsque les résultats sont certains[5]. C’est-à-dire qu’une action dont les résultats sont certains n’a pas besoin de reposer sur la confiance. Celle-ci ne rentre d’ailleurs plus tout en jeu puisqu’elle n’est existante que dans un environnement incertain. En revanche, la confiance est un facteur important dans les environnements à hauts risques avec de grandes incertitudes dans lesquels les décisions doivent être prises rapidement à partir d’informations imparfaites. En effet, celles-ci rendent l’environnement plus complexe à appréhender. Dans de telles situations, la tendance d’un individu à faire confiance est déterminée par les contextes individuels, culturels et organisationnels (Lee et See [66]).

Ces deux derniers éléments de contexte influencent la confiance lorsque des individus interagissent, au sens où ils s’informent de la fiabilité des autres (par exemple, leurs compétences). Cette influence existe aussi à travers l’établissement de normes génériques et de comportements prédictibles [90]. Quant au contexte individuel, il réfère à l’historique des interactions avec le fiduciaire et à l’inclination à faire naturellement confiance, variable d’un individu à l’autre.

Ces variations individuelles ne sont pas négligeables. En effet, elles affectent la fiabilité technique des automates, sans être reliées aux caractéristiques de ces derniers. Des recherches ont montré que les tendances à la confiance en tant que trait de personnalité peuvent être mesurées et que ces tendances influencent le comportement de manière prédictible [82]. Cela n’empêche pas Lee et *al.* de recommander la prise en compte de ce facteur lors de la conception du système.

[4] Trust guides - but dœs not completely determine - reliance
[5] Premier critère de la confiance, Bhattacharya

2.2.1.2 Prise de décision

Dans certaines situations, les acteurs connaissent le résultat de l'action d'autrui avant de devoir agir, tandis que dans d'autres cas, ils doivent agir simultanément ou sans connaissance des conséquences des actions d'autrui. Dans ces derniers cas, les individus prédiront les actions des autres et y assigneront une probabilité subjective [94]. Dans de telles situations et avec la notion de risque, la confiance devient un calcul de risque, tel que le conçoit Dasgupta [38]. Nous parlons ici de confiance décidée qui fait suite à une réflexion. Avec cette vision de la confiance, cette réflexion prend alors la forme d'une évaluation de risque dont le résultat déterminera le degré de confiance. Celle-ci repose sur des biais de décision ainsi que des heuristiques décrits par Tversky et Kahnemann [111]. En effet, selon Luhmann, une décision de confiance présuppose un risque, et la confiance, en plus de réduire la complexité de l'environnement, représente un degré d'acceptation du risque. Dasgupta [38] considère le risque comme le calcul de la probabilité d'évènements rares aux conséquences négatives. On peut alors considérer le degré d'acceptation du risque comme un seuil au-dessus du quel le risque devient trop important pour faire confiance.

$$P(\text{risque})_{encouru} < P(\text{risque})_{accepte} \Rightarrow \text{décision}(\text{confiance}) \tag{2.1}$$

$$P(\text{risque})_{encouru} > P(\text{risque})_{accepte} \Rightarrow \text{décision}(\text{méfiance}) \tag{2.2}$$

Pour Bhattacharya et *al.* [7], la probabilité assignée représente la force d'une attente, ou le degré de confiance dans la prédictibilité du fiduciaire. Une grande prédictibilité est un facteur clé pour favoriser la confiance, en permettant au mandataire d'établir des attentes précises face aux capacités et au comportement du fiduciaire.

Cette notion de force d'attente se retrouve chez Sutcliffe [107], qui décide de déléguer des tâches par confiance si la prise de risque est moindre que le risque maximum toléré. On peut le formuler directement en termes de confiance en considérant que si la confiance envers le fiduciaire est supérieure à une confiance minimale requise alors on peut déléguer en confiance une tâche à celui-ci.

$$\text{Confiance}_{reelle} > \text{Confiance}_{minimale\ requise} \Rightarrow \text{déléguer}(\text{tâche}) \tag{2.3}$$

2.2.1.3 Résultats et conséquences

Selon Rempel [94], la confiance se définit par : *"l'attente relative à une probabilité subjective qu'un individu attribue à l'occurrence d'un ensemble d'évènements futurs"*[6]. Lee et See reformulent l'idée en disant : *"la confiance concerne une attente*

[6]Expectation related to subjective probability an individual assigns to the occurence of some set of future events

ou une attitude dépendant de la vraisemblance d'une réponse favorable"[7]. Un individu va, en plus de prédire l'action d'autrui, assigner une probabilité représentative de l'efficacité de cette individu à l'atteinte du résultat escompté. Cette prédiction est influencée par l'importance qu'attache cet individu à atteindre un résultat spécifique, souvent lié aux conséquences d'une réussite ou d'un échec. Dans l'approche décisionnelle de la confiance que propose Sutcliffe [107], ce facteur d'importance va largement déterminer, chez le décideur, la confiance minimum requise pour décider s'il peut faire confiance au fiduciaire.

Le fait de réussir ou d'échouer à atteindre un résultat favorable affecte la confiance future de l'individu en créant, renforçant ou réfutant les attentes sur les compétences du fiduciaire [7]. Pour en revenir aux travaux de Sutcliffe, ce mécanisme est modélisé par la prise en compte des transactions passées pour l'évaluation de la confiance, en vue d'une nouvelle transaction. Par contre dans une nouvelle relation, la confiance ne peut reposer sur cet historique — puisque inexistant. Il est alors nécessaire pour le mandataire de s'informer sur le fiduciaire afin de formuler une première opinion.

2.2.1.4 Information sur le fiduciaire

Les diverses et nombreuses recherches qui ont été conduites dans le domaine de la confiance ont généré une grande quantité de définitions. Nombre de ces définitions se focalisent sur la confiance en tant qu'attitude ou attente qu'un opérateur humain a envers l'automate. Le consensus qui en résulte est de définir la confiance comme : "l'attente de la réalisation de un rôle de façon compétente sur le plan technique"[8] [6], ou "les attentes des obligations et des responsabilités du fiduciaire, c'est alors l'attente que certains individus dans nos relations sociales aient des obligations morales et la responsabilité de démontrer un intérêt spécial pour les intérêts des autres avant les leurs"[9] [6]. Bhattacharya et *al.* [7] décrivent la confiance comme une construction multidimensionnelle basée sur les caractéristiques du fiduciaire. Les caractéristiques pertinentes concernent celles informant de l'habileté de l'individu à achever le but du mandataire.

Lee et See affirment que la confiance évolue dans un contexte individuel, culturel et organisationnel. Or, à l'intérieur de ce contexte, le rôle des attentes est important. Par exemple, les individus suivent une approche d'apprentissage social dans laquelle les attentes pour une situation particulière sont déterminées par des expériences précédentes, spécifiques à des situations perçues comme similaires. La confiance est alors une attente générale qui est indépendante d'expériences spécifiques, mais qui est plutôt basée sur la généralisation d'un large nombre d'expériences diverses [66]. Par exemple, le contexte organisationnel reflète l'interaction entre les gens, ce qui

[7]Trust concerns an expectancy or an attitude regarding the likelihood of favorable responses

[8]Expectation of technically competent role performance

[9]Expectations of fiduciary obligation et responsibility, that is, the expectation that some others in our social relationships have moral obligations and responsibility to demonstrate a special concern for others' interests above their own

les informe de leurs fiabilités respectives, pouvant inclure réputations et rumeurs. Quant au contexte social, il est composé d'un ensemble de normes et d'attentes. Par conséquent, les individus commence à interagir entre eux ou avec des automates, avec un degré de confiance prédéterminé et basé sur ces contextes. Ainsi, Lee et See [66] affirment que les individus seront initialement confiants envers un ingénieur, non pas à cause de l'habileté spécifique de la personne, mais sa qualification. En d'autres termes, ils font confiance au métier et non à la personne.

Parallèlement, Muir déclare que les décisions d'un opérateur humain pour autoriser (respectivement annuler) un automate à contrôler un processus démontrent un certain niveau de confiance envers l'automate. Le modèle présenté dans ces recherches définit la confiance en convergence avec Lee et See, et l'étend par une description multidimensionnelle du caractère de la confiance, incluant trois bases spécifiques (ou attentes) : compétence technique, persistance, et responsabilité fiduciaire - issues des travaux de Barber [6]. Ces bases sont similaires aux concepts décrit par Schoorman *et al.* [99], à savoir l'habileté, l'intégrité et la bienfaisance.

Compétence technique - habileté :

La compétence technique signifie simplement que l'on croit que le fiduciaire aura la capacité de réaliser correctement la tâche qui lui est assignée [80]. Lee et See définissent l'habileté comme "le groupe des aptitudes, des compétences et des caractéristiques que permet au fiduciaire d'agir dans le domaine"[10] ce qui rejoint la définition qu'en donne Mayer *et al* [72]. La perception que nous avons de l'habileté du fiduciaire, ainsi que des caractéristiques telles que la fiabilité et la prédictibilité, contribue à une attente en termes de performance, qui est un composant principal de la confiance et tout spécialement dans les premières étapes de la relation de confiance. La section 2.2.2 étudie la relation entre performance et confiance.

Persistance - intégrité :

La persistance réfère à l'immuabilité d'une attente qui, par exemple, permet à l'opérateur de construire une représentation mentale de l'automate à travers des expériences d'interactions avec celui-ci [80]. Ceci est similaire à l'intégrité que Lee et See [66] définissent comme "le degré d'adhésion du fiduciaire à un ensemble de principes que le mandataire considère acceptable"[11]. La perception de l'intégrité de l'individu de confiance peut être considérée comme l'observation de ses capacités sur une longue période et nous offre une attente de fiabilité qui participe à la dynamique de la confiance décrite en section 2.2.2. Mayer *et al.* [72] dans leur définition de l'intégrité précisent qu'il s'agit de l'adhésion perçue par le mandataire. Le mandataire évalue sa confiance en se basant sur ces perceptions et donc la représentation qu'il se fait du fiduciaire.

Responsabilité fiduciaire - bienfaisance :

La bienfaisance est définie par Schoorman *et al.* [99] comme la volonté du fiduciaire à vouloir bien faire pour le mandataire. Elle dépend du degré de responsabilité

[10] The group of skills, competencies, and characteristics that enable the trustee to influence the domain

[11] The degree to which the trustee adheres to a set of principles the trustor finds acceptable

du fiduciaire qui s'applique lorsque les compétences de celui-ci sont inconnues et que l'on doit présupposer qu'il agira. Ceci arrive souvent lorsque l'opérateur présuppose qu'un système atteindra ses critères de performances théoriques s'il est opéré correctement [80]. Cette responsabilité fiduciaire est similaire, chez Lee et See [66], au terme de bienfaisance, qu'ils définissent comme "l'importance avec laquelle les intentions et les motivations du fiduciaire sont alignées avec celles du mandataire"[12]. Dans les relations avancées de la confiance, la perception qu'a le mandataire de la bienfaisance du fiduciaire tend vers une relation dénuée de tout doute. C'est à dire qu'il se fie au fiduciaire même si des actions spécifiques et leurs conséquences peuvent être inconnues dans l'instant.

2.2.2 Dynamique de la confiance

Mayer *et al.* [72] considèrent que la dimension temporelle a un impact sur le lien entre habileté, intégrité et bienfaisance. En effet, ils formulent l'hypothèse que dans une relation nouvelle, l'habileté et l'intégrité seront prépondérantes lorsque le mandataire déterminera son degré de confiance. Tandis que la part de confiance liée à la bienfaisance prendra une place de plus en plus conséquente avec le temps. Sans remettre en cause ces deux propositions, Schoorman *et al.* [99] remettent en doute l'indépendance entre les notions d'intégrité et de bienfaisance en raison d'une forte corrélation entre ces deux variables. S'ils ne remettent en doute ce modèle, c'est parce qu'ils considèrent que les relations de confiance n'ont pas duré suffisamment pour laisser le temps agir sur la bienfaisance.

Par ailleurs Rempel et *al.* [94] définit trois dynamiques dans la confiance : prédictibilité, fiabilité et foi (absence de doute). Ces dynamiques apparaissent consistantes avec les facteurs de performance, de processus et d'usage définis par Lee et See [66]. Chacune de ces dynamiques s'applique plus ou moins en fonction du degré de maturité de la relation de confiance, et chacune requiert des quantités et des types variables d'informations afin de soutenir une confiance appropriée. Cela rejoint les propositions de Mayer *et al.* sur la prépondérance de certains facteurs de la confiance par rapport à d'autres en fonction du temps.

2.2.2.1 Prédictibilité

La prédictibilité est importante durant les premières étapes d'une relation de confiance, et requiert une visibilité sur les comportements spécifiques ou sur les performances du systèmes. Par exemple, des opérateurs jugeront l'automate sur ses capacités à délivrer les résultats désirés. Cependant, plusieurs facteurs influencent le jugement de prédictibilité.

Le premier est la performance réelle de l'automate. Lee et See [66] décrivent la performance comme "ce que l'automate fait. Plus spécifiquement, la performance se

[12] the extent to which the intents and motivations of the trustee are aligned with those of the trustor

réfère à la compétence ou l'expertise qui a été démontrée par l'habilitée à atteindre les buts de l'opérateur"[13].

Le deuxième facteur est la prédictibilité perçue (ou transparence). L'opérateur doit être capable d'observer le comportement de l'automate afin d'estimer sa prédictibilité. En général, un automate simple sera plus prédictible et observable qu'un automate qui a de nombreux degrés de libertés. Si l'on se réfère à la taxonomie des degrés d'autonomie de HACT [12] (présentée à la section 1.3), on se rend compte qu'un certain nombre de niveaux d'autonomie ne permettront pas cette observation. En effet, le paramètre de transparence fonctionnelle détermine la visibilité de l'opérateur sur le système. Ainsi, s'il est partiellement transparent l'opérateur ne pourra que difficilement se construire un modèle mental du système et donc déterminer sa prédictibilité. Cette construction passe par l'interaction que l'opérateur a avec son système et de la perception qu'il a de la situation courante. Ceci est parfaitement intégré au sein du modèle de partage d'autorité défini par Saget *et al.* [97] (introduit à la section 1.4.2.1).

Enfin, la stabilité de l'environnement peut affecter la prédictibilité [80]. Un système parfaitement prédictible dans un environnement stable peut, si il est sensible à son environnement, devenir imprédictible lorsque l'environnement perd sa stabilité. C'est ce qu'on nomme l'imprédictibilité apparente en opposition à l'imprédictibilité inhérente qui est intrinsèque au système. Dans ce dernier cas, le système reste imprédictible au sein d'un environnement stable. La prédictibilité perçue va jouer un rôle important dans le discernement du type d'imprédictibilité auquel fait face l'opérateur.

Durant le développement et les tests opérationnels d'un système, les testeurs mesurent attentivement la performance des composantes du système et conduisent des essais contrôlés et répétés afin d'isoler les différentes variables possibles. Ceci a pour but de minimiser l'imprédictibilité apparente. Cependant, appliquer toutes les variables est impossible dans un environnement opérationnel. Les opérateurs doivent se fier à la transparence du système, à leurs entraînements et à leur expérience pour prédire précisément le comportement du système.

2.2.2.2 Fiabilité - *Dependability*

Lorsque la relation de confiance est mature, l'attention change, passant de comportements spécifiques à une évaluation d'une fiabilité générale. Des opérateurs familiers avec un automate commenceront à faire confiance, du fait de leur expérience globale plutôt que de comportements spécifiques, et auront une plus grande compréhension des situations dans lesquelles l'automate peut être indigne de confiance. Lee et See [66] décrivent cette tendance en termes de processus, qui est "le degré auquel les algorithmes d'un automate sont appropriés à la situation et sont capables

[13] What the automation dœs. More specifically, performance refers to the competency or expertise as demonstrated by its ability to achieve the operator's goals

de remplir les objectifs de l'opérateur"[14]. L'opérateur ne porte plus un regard au cas par cas, mais appréhende les compétences du système en dehors de tout contexte. Autrement dit, l'opérateur ne s'intéresse pas uniquement au fait accompli par le système mais à l'élaboration de règles qui détermine le domaine de compétence de celui-ci. Si on se réfère au paradigme de Rasmussen, on peut considérer que l'opérateur est monté d'un niveau dans sa perception du système. Sa confiance n'est plus définie en réaction à chaque nouvelle situation, mais par l'application de règles plus générales.

Le jugement d'un opérateur sur la fiabilité de l'automate est amélioré lorsque l'opérateur place l'automate dans des situations incertaines et risquées, lui permettant d'observer comment l'automate réagit en dehors de son fonctionnement nominal [80] (c'est-à-dire en dehors de son cadre de fonctionnement théorique, pour lequel il a été conçu). Ceci se réalise durant les tests et les évaluations ou durant l'entraînement, mais peut aussi avoir lieu dans un environnement opérationnel. Pour finir, Lee et See [66] affirment que "les opérateurs feront confiance à un automate si ses algorithmes sont compréhensibles et capables d'atteindre les objectifs de l'opérateur dans la situation courante"[15]. Cette affirmation est renforcée par les inférences que l'opérateur génère en observant les performances de l'automate durant une période temporelle donnée.

2.2.2.3 Foi

Dans les dernières étapes, la foi représente une relation de confiance complètement mature. Ici, la foi est définie comme "la fermeture au doute face à un avenir incertain"[16]. Dans cette étape, la perception qu'a l'opérateur de la pertinence et de la flexibilité de l'automate lui permet de contrôler le système efficacement, même sans une complète compréhension des comportements complexes de l'automate. La foi, dans ce contexte, est similaire à l'usage que Lee et See [66] définissent comme "Le degré auquel l'automate est utilisé dans le domaine de fonctionnement défini par ses concepteurs"[17]. Ceci suppose que les intentions du concepteur ont été communiquées à l'opérateur. Cependant, l'opérateur prend aussi conscience du potentiel de l'automate pour des interactions imprévues, ainsi que des limites inhérentes au système au cours de circonstances où les procédures sont dégradées [80].

Il est à noter que dans la dynamique de la confiance selon Mayer *et al.* [72], cette troisième étape n'existe pas. En effet, pour eux la confiance est principalement déterminée par la prédictibilité puis, progressivement et avec le temps, elle est déterminée par la fiabilité. Une fois la fiabilité prépondérante, il n'y a pas de troisième

[14] The degree to which the automation's algorithms are appropriate for the situation and able to achieve the operator's goals

[15] The operator will tend to trust the automation if its algorithms can be understood and seem capable of achieving the operator's goals in the current situation

[16] *Closure against doubt in the face of an uncertain future*

[17] *The degree to which the automation is being used within the realm of the designer's intent*

étape dans la relation de confiance. Rempel pour les relations interpersonnelles et Muir pour les relations Homme-machine considèrent cette dernière étape comme un “saut de la foi”[18], c’est-à-dire une étape où le mandataire malgré ses lacunes dans la compréhension du système décide de lui faire confiance. Il est d’ailleurs très important à ce moment que l’opérateur n’ait pas une confiance excessive envers le système.

2.3 Modèle de la confiance

Jusqu’à présent nous avons abordé les différentes composantes de la confiance et son évolution au cours du temps. Nous allons à présent nous intéresser à l’intégration de ces éléments au sein de plusieurs modèles de la confiance. Ceux choisis sont représentatifs de différents besoins : la prise de décision, la prédiction ou simplement la description.

2.3.1 Modèle de Sutcliffe - décision

Le premier modèle que nous abordons est conçu pour la modélisation des décisions de confiance. Dans le domaine des systèmes multi-agents, un certain nombre de recherches est conduit pour la modélisation de la confiance entre agents. On retrouve ainsi les travaux de Marsh [71] qui définit un certain nombre de formalismes mathématiques, ainsi que ceux de Castelfranchi *et al.* [13, 14, 15] qui utilisent une architecture BDI[19] pour définir la confiance au sein d’un agent et qui prennent en compte les modèles sociaux de la confiance. Enfin, les travaux de Sutcliffe [107], que nous allons développer ici, se différencient de ceux de Castelfranchi *et al.* par leurs applications à des cas d’études. En effet, les travaux de Castelfranchi et al. bien qu’ils soient le fruit d’une synthèse de la littérature, ne restent qu’un travail théorique qui n’a pas été éprouvé expérimentalement.

Sutcliffe considère que la confiance se place dans le cadre de la prise de décision. Afin de modéliser la confiance, elle part alors de l’affirmation que les décisions sont influencées par des évènements et une situation à un moment donnée combinés à aux a priori et aux attitudes d’un individu vis-à-vis du problème de décision. Elle décrit alors le processus de décision de la confiance en quatre étapes qui prennent compte de la réputation et des expériences antérieures 2.1.

La première étape est l’évaluation de la nécessité de la confiance. En effet, nous avons vu auparavant que la confiance présuppose une situation de risque et de vulnérabilité. Ainsi, lors de cette première étape le mandataire — qui initialise la relation avec le fiduciaire — évalue une partie des critères que nous avons abordé précédemment. Le contexte ainsi que les risques sont évalués. Le résultat de cette analyse va déterminer un seuil de décision pour la confiance.

[18] *leap of faith*
[19] *belief, desire, intention*, BDI

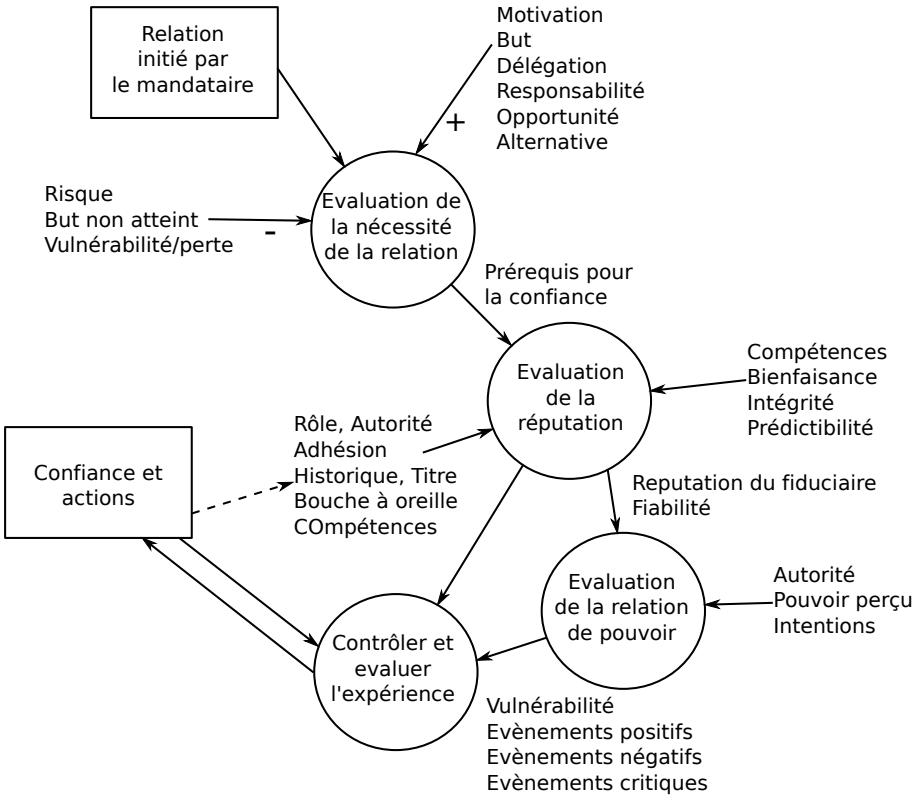

FIG. 2.1 : Modèle de la confiance par Sutcliffe [107].

Lors de la deuxième étape, le mandataire va cette fois analyser les informations qu'il a du fiduciaire. En effet, son degré de confiance vis-à-vis de celui-ci est déterminé la perception qu'il a de son habileté, de sa bienfaisance et de son intégrité. Notons que l'on retrouve ici les trois dimensions de la confiance telle que définie par Schoorman et al. [99]. Cette évaluation se base la première fois essentiellement sur la réputation — c'est-à-dire des informations de seconde main. Par la suite, au fur et à mesure des interactions entre le mandataire et le fiduciaire, cette évaluation pourra se baser uniquement sur l'expérience acquise des échanges précédents. Cela signifie que la confiance est alors évaluée par le mandataire à partir des interactions qu'il a eu avec le fiduciaire.

La troisième étape consiste en une évaluation du rapport de force d'un point de vue autoritaire entre le mandataire et le fiduciaire. Elle consiste a évalué les intentions du fiduciaire pour voir si ce dernier sera ou non coopératif.

Enfin la dernière étape, une tâche a été déléguée, et le mandataire n'a plus qu'à surveiller son déroulement et à se faire sa propre idée du fiduciaire. C'est-à-dire se construire sa propre représentation du fiduciaire, au lieu de se baser sur sa réputation. Il y a ici une interaction possible entre les deux partis. Avec un agent humain cela peut se traduire par la mise en œuvre d'un dialogue. En effet, afin de

surveiller le fiduciaire, le mandataire doit s'informer de ce qu'il fait. Pour cela il peut l'observer mais aussi interagir avec celui-ci afin d'inférer un modèle de celui-ci. Et ce sont par ailleurs les informations acquises lors de ces échanges qui détermineront l'évolution de la confiance pour les décisions suivantes.

2.3.2 Modèle de Muir - prédiction

Le modèle que nous abordons maintenant est un modèle prédictif de la confiance. En effet, il a pour but de modélisation la confiance d'une opérateur en se basant sur les mêmes observables que l'opérateur. Il simule donc le processus cognitif d'évaluation de la confiance afin de prédire qu'elle sera la confiance de l'opérateur au vu des observables d'entrées.

Ce modèle est fondé sur les travaux de deux auteurs : d'une part la vision multidimensionnelle de la confiance définie par Barber [6] à savoir les compétences, la persistance et la responsabilité ; et d'autre part les trois dynamiques de la confiance définies par Rempel [94] que sont la prédictibilité, la fiabilité et la foi. Muir considère que ces deux visions sont orthogonales et que la confiance repose sur une évaluation de chacun des croisements entre ces deux représentations (par exemple l'association compétences et fiabilité).

D'un point de vue plus formel, le modèle de Muir décrit le processus d'évaluation de la confiance d'un opérateur i vis-à-vis d'un automate j (une fonction précise dans un mode opératoire donné, c'est-à-dire configuré à un niveau d'autonomie donné). Nous avons j qui possède des compétences, des responsabilités et certains attributs et qui va avoir un comportement pour lequel l'opérateur i attribue un certain degré de confiance (fig.2.2). Celui-ci est déterminé selon les trois dimensions de la confiance (Barber [6]) ainsi que ses trois dynamiques (Rempel [94]) :

- Les attentes de i qui concernent la persistance du monde basé sur :
  - Prédictibilité : les évènements sont conformes aux lois naturelles (*e.g.* lois de la physique) ;
  - Fiabilité : assujettissement aux lois naturelles ;
  - Foi : les lois naturelles sont constantes.
- Les attentes de i qui concernent les compétences du système basées sur :
  - Prédictibilité : le comportement de j est prédictible ;
  - Fiabilité : j démontre ses compétences ;
  - Foi : j continuera de démontrer ses compétences à l'avenir ;
- Les attentes de i qui concernent les responsabilités du système basées sur :
  - Prédictibilité : le comportement de j est consistant à ses responsabilités (*i.g.* la réalisation de ses objectifs) ;
  - Fiabilité : j assume ses responsabilités ;
  - Foi : j continuera d'assumer ses responsabilités à l'avenir.

Le degré de confiance est alors une combinaison de ces attentes.

Ce modèle évalue la confiance sur un certain nombre de critère qui repose tous sur la perception du mandataire. Ces dernières reposent sur les interactions entre

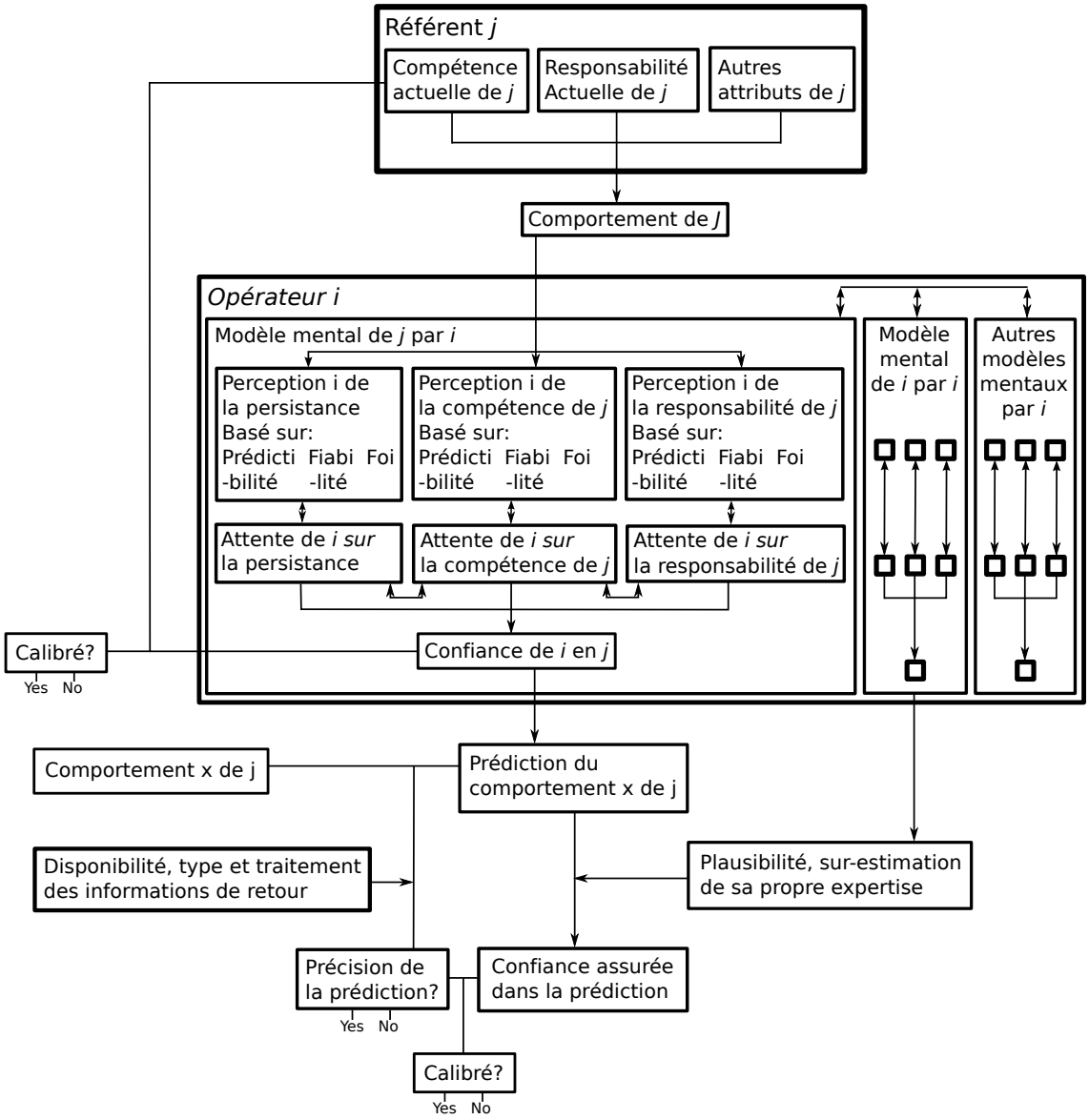

FIG. 2.2 : Modèle de la confiance par Muir [80].

l'opérateur et son système. De plus, dans une expérimentation, Muir montre, à l'aide de son modèle, que les décisions d'un opérateur pour autoriser (respectivement empêcher) un automate à contrôler un processus démontrent un certain niveau de confiance envers le système. Ainsi, si l'on veut prédire l'usage de l'automate par un opérateur en fonction de la confiance que ce dernier lui accorde, il est nécessaire d'avoir un modèle de la confiance au sein des interactions Homme-machine.

2.3.3 Modèle de Lee - description

Lee et See dans leur état de l'art sur la confiance en l'automatisation synthétisent l'ensemble des points précédemment abordés [66]. De cette analyse, ils déduisent un modèle descriptif de la confiance (fig.2.3).

La confiance représente pour eux une intégration de l'expérience du couple Homme/automate. Ainsi la confiance détermine l'usage qui est fait de l'automate. Les échecs et réussites résultants auront pour effet de faire fluctuer la confiance. L'interaction des deux parties (mandataire et fiduciaire) est donc un élément central de ce modèle. En effet, c'est au travers du dialogue Homme-machine que s'établit le retour d'expérience et donc son intégration. C'est donc grâce à cette interaction que le modèle forme une boucle dynamique à l'origine de l'évolution de la confiance.

Cette boucle dynamique se décompose en six étapes. Quatre déterminent les actions de l'opérateur, et les deux autres forment la boucle de retour comme illustré fig.2.3 :

Assimilation des informations et formation des connaissances : lors des premières interactions entre un opérateur et un système, les informations et les connaissances qui vont influer sur la confiance vis-à-vis du système sont fortement liées à la réputation du système (ex : les retours d'expériences issus d'autres opérateurs) et à la formation que l'opérateur a suivi. Avec le temps l'opérateur formera ses propres connaissances et son propre avis à propos du système, en revanche cette acquisition d'informations est alors sujette aux capacités de l'interface Homme-machine qu'il aura à sa disposition. En effet si l'on reprend la taxonomie de HACT [12], certains paramètres déterminent la transparence fonctionnelle du système, la transparence de l'information et l'interactivité et ont pour conséquent une perception différente du système par l'opérateur. On retrouve donc ici l'ensemble des informations qui portent sur le fiduciaire.

Évolution de la confiance : celle-ci est pilotée par les connaissances de l'opérateur et sa perception du système. Cette évolution peut être influencée par des critères sociaux-culturels mais aussi individuels. Par exemple, certain individus sont plus disposés à accorder leur confiance que d'autres.

Formation des intentions : dans la formulation des intentions nous retrouvons la notion de décision. Ainsi, un grand nombre de paramètre environnementaux comme le risque perçu, la confiance en soi ou encore la charge cognitive sont évalués. C'est à ce niveau que l'on retrouve l'ensemble des facteurs de la confiance à l'exception des informations sur le fiduciaire.

Action de confiance : ici les intentions sont traduites en action, comme nous l'avons vu plus tôt chez Cohen [26], les contraintes temporelles jouent un rôle important sur la décision de l'opérateur ;

Automate : l'automate réalise les tâches demandées par l'opérateur. Il peut alors être sujet à des facteurs environnementaux qui affectent ses capacités (ex : une

panne de courant) ;

Retour d'informations : ils sont fournis par le système et peuvent être détaillés selon plusieurs niveaux de détails (*e.g.* système, fonction, sous-fonction), être sous forme diverses (*e.g.* données brutes, graphismes). Ces informations seront ensuite assimilées par l'opérateur, ce qui correspond à la première étape de ce modèle ;

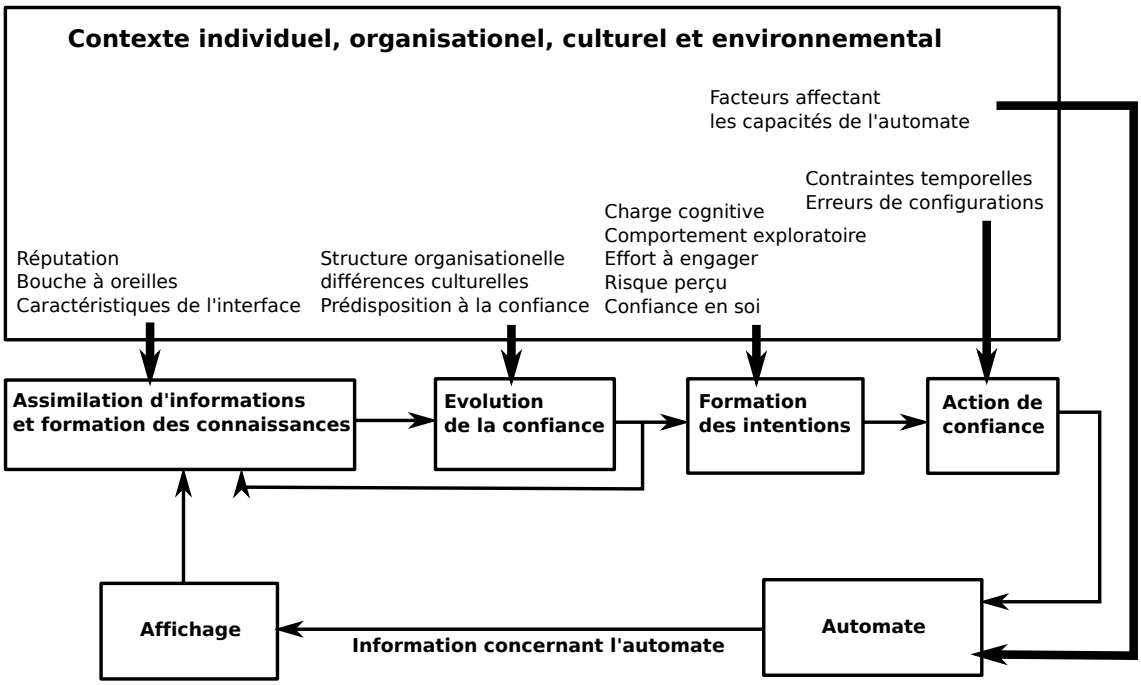

FIG. 2.3 : Modèle descriptif de la confiance par Lee [66].

2.4 DISCUSSION

Les modèles vus précédemment sont très différents les uns des autres. Selon que l'on souhaite prendre une décision, estimer la confiance ou simplement la décrire, le modèle à choisir sera différent.

Ainsi, le modèle de Lee permet une description exhaustive de la confiance dans l'automatisation. Toutefois, s'il dispense un regard critique pour la conception des systèmes, il ne donne aucun outil permettant son intégration au sein du système lui-même. Pour cela, il faut s'intéresser davantage au modèle de Muir.

Le modèle de Muir perd beaucoup en exhaustivité comparativement au modèle de Lee. Cela lui permet en revanche son intégration au sein d'un système [31]. Ainsi il est donné la possibilité au système d'estimer la confiance de l'opérateur et de s'adapter en conséquence.

Quant au modèle de Sutcliffe, il se démarque complètement des travaux de Lee et Muir. En effet, il s'intéresse à la question "puis je faire confiance ?" tandis que Lee et Muir cherchent à estimer la confiance et son évolution. Nous passons donc d'une modélisation de la confiance dans le temps (Lee et Muir) à une modélisation d'une décision basée sur la confiance (Sutcliffe).

Enfin, nous avons montré qu'au sein de tout ces modèles, l'interaction joue un rôle important. En effet, c'est au travers de celle-ci que le mandataire acquiert des informations sur le fiduciaire et qu'il s'en construit un modèle mental. Or l'on constate que bien que l'interaction soit centrale, elle n'est pas exploitée dans ces modèles. Leur modélisation de la confiance est fondée sur ses caractéristiques alors que l'interaction pourrait être significative dans son évaluation. Il ne faut pas oublier que si l'on aborde la modélisation de la confiance, c'est pour pouvoir prédire et anticiper des situations où l'opérateur serait trop ou pas assez confiant. Or pour cela nous voulons la mesurer.

2.5 Evaluation de la confiance

Les approches actuelles se placent systématiquement dans une optique de simulation du processus de confiance pour estimer quelle est la confiance donnée ou quelle va être la confiance de l'opérateur. Pour cela, des expériences ont été menées afin de paramétrer ces modèles. Ainsi Muir [81] évalue à la fois la confiance et ses facteurs (compétence, prédictibilité, etc.) à l'aide de questionnaires subjectifs. Les résultats de ces expériences aboutissent à des conseils sur la formation des opérateurs et sur la conception des systèmes. En effet, le modèle qui repose sur des observables subjectifs tel que les compétences perçues, la prédictibilité perçue, est dépendant des questionnaires correspondants.

2.5.1 Questionnaire d'évaluation

2.5.1.1 Questionnaire de Jian

Nous avons présenté précédemment un certain nombre de modèle de la confiance, mais ces derniers n'abordaient pas directement la problématique de l'évaluation de la confiance. Par contre, à des fins de validations, des questionnaires d'évaluation de la confiance sont utilisés. Il est a noté qu'un état de l'art sur l'évaluation de la confiance nous conduit principalement sur ces derniers.

Ces questionnaires sont nombreux et différents selon qu'ils ciblent une relation de confiance Homme-Homme [73] ou Homme-machine [81, 65].

Des travaux menés par Jian ont permis la création d'un questionnaire empirique qui adresse à la fois la confiance Homme-Homme et Homme-machine. Il [54] établit trois constats au vu de l'état de l'art sur l'évaluation de la confiance :

- Les questionnaires (antérieurs à ces travaux) reposent sur des aspects théoriques de la confiance et aucun n'est basé sur une analyse empirique. En effet

si nous regardons le questionnaire établi par Muir [81] pour valider son modèle, celui-ci reprend les éléments clés de son modèle (prédictibilité, compétences, *etc.*) auxquels il adjoint une échelle continue (de 10cm) avec les labels "pas du tout" et "extrêmement haut" à chaque extrémité. Afin de clarifier le questionnaire, des définitions sont adjointes à chaque des mots clés (e.g. "à quel point le système fonctionne correctement ?"). Il demande aussi aux sujets d'évaluer leur confiance, ce qui lui permet d'établir les relations entre les dimensions de la confiance. Dans les relations Homme-machine, les questionnaires ont souvent pour but de confronter les éléments d'une hypothèse à une évaluation directe de la confiance. Ainsi Lee et Moray, dans une première expérience, réalisent un questionnaire équivalent à celui de Muir. Tandis que dans une seconde expérience, où ils veulent étudier les liens entre confiance et confiance en soi, Lee et Moray réduise leur questionnaire à deux types de questions : celles qui portent sur la confiance du sujet vis-à-vis d'un automate, et celles qui portent la confiance en soi de l'opérateur à réaliser la tâche de cet automate. Enfin un dernier exemple dans le cadre des relations interpersonnelles, le questionnaire de McAllister [73] est lui basé sur une pré-sélection de 40 questions choisis par l'auteur. Ces choix sont soumis à une communauté d'experts afin d'être classés selon qu'ils se rapportent à une confiance cognitive, une confiance émotionnelle, les deux ou aucun des deux. A partir de cette classification McAllister réduit son questionnaire à onze questions. Ce dernier questionnaire est un peu plus empirique mais se cantonne à une présélection qui, elle, repose sur des aspects théoriques.

- Les questionnaires se basent sur l'hypothèse que confiance et méfiance sont opposées.
- Les questionnaires se basent sur l'hypothèse que les notions de confiance Homme-Homme peuvent s'appliquer aux relations Homme-machine.

Jian propose donc de conduire une expérience en plusieurs phases afin d'établir une échelle empirique de la confiance mais aussi pour évaluer la similarité des relations Homme-Homme et des relations Homme-machine. Cette expérience consiste en une étude linguistique (en anglais) qui se divise de la manière suivante :

- collecte du vocabulaire lié à la confiance à l'aide de dissertation sur la confiance demandée à des étudiants issus de formations littéraires (linguistique ou anglais),
- questionnaire d'étude pour évaluer la distance des mots recueillis aux notions de confiance et de méfiance afin d'établir l'existence d'une opposition entre ces deux notions et de déterminer si les concepts de confiance, de confiance entre individus et de confiance Homme-machine sont similaires. En effet on pourrait considérer que le comportement vis-à-vis d'un individu, d'un machine, ou de quoique ce soit d'autres (e.g. un animal) puissent être différent. Ces différences peuvent alors s'insinuer dans la conception de la confiance et ainsi créer une différenciation selon que l'on interagit avec un individu, une machine ou autres choses.

– comparaison de similarité entre des paires de mots afin d'établir une échelle multi-dimensionnelle de l'évaluation de la confiance.

Jian montre à l'aide de cette expérience que les concepts de confiance, de confiance Homme-Homme et de confiance Homme-machine font référence à un vocabulaire similaire. Il est donc possible d'établir un même et unique questionnaire pour chacune de ces situations. Il montre aussi que la confiance Homme-Homme est moins marquée que pour les deux autres. Autrement dit, un même mot aura une connotation de confiance (méfiance) plus positive (négative) dans le cadre d'une confiance Homme-machine que pour une confiance Homme-Homme. Il propose aussi en conclusion de son étude un questionnaire de la confiance en douze points[20] :

1. Le système est trompeur
2. Le système se comporte de manière sournoise
3. Je me méfie des intentions, des actions, et des sorties du système
4. Je suis vigilant vis-à-vis du système
5. Les actions du système ont une issue nuisible ou préjudiciable
6. Je suis confiant envers le système
7. Le système offre une sécurité
8. Le système est intègre (honnête)
9. On peut compter sur le système
10. Le système est digne de confiance
11. Je peux faire confiance au système
12. Je connais bien le système

On retrouve ici le vocabulaire présenté en introduction de ce chapitre :

[20]

1. The system is deceptive
2. The system behaves in an underhanded manner
3. I am suspicious of the system's intent, action, or outputs
4. I am wary of the system
5. The system's actions will have a harmful or injurious outcome
6. I am confident in the system
7. The system provides security
8. The system has integrity
9. The system is dependable
10. The system is reliable
11. I can trust the system
12. I am familiar with the system

– le mot “confident” qui se réfère au nom “confidence”, ainsi que le mot “trust” ont été traduit par le mot “confiance”. La distinction confiance assurée et confiance décidée n’a pas été maintenue dans ce questionnaire afin de facilité sa compréhension. De plus ces deux mots sont employés dans des phrases aux structures suffisamment différentes pour qu’elles ne perdent pas leurs sens distinctifs malgré cette simplification.
– le mot “dependable” fait référence à “dependability” dont le sens est la possibilité de déléguer une tâche. Nous avons donc traduit l’adjectif par l’expression “compter sur quelqu’un”.
– le mot “reliable” est l’adjectif associé à “reliability” qui se définit par la fiabilité technique. On la traduit par l’expression “digne de confiance” pour que l’expression garde son sens aussi dans les relations interpersonnelles.

Enfin, des travaux plus récents ont été réalisés par une autre équipe de chercheurs afin de valider ce questionnaire à l’aide une analyse factorielle confirmatoire[21] [105].

2.5.1.2 Utilisation du questionnaire

Ce questionnaire a été réutilisé pour évaluer la confiance dans plusieurs expériences. L’exploitation de ce questionnaire varie selon les travaux. Ainsi certains auteurs reprennent le questionnaire à l’identique pour l’évaluation de la confiance. C’est le cas de Jiang *et al.* [55] qui étudie l’impact de la confiance sur un système automatique d’inspection de la qualité, ou encore de Cramer *et al.* [29] sur la confiance des utilisateurs d’un système anti-spam. Cramer *et al.* dans une étude antérieure [28] n’avaient repris que partiellement le questionnaire en ne conservant que les questions portant sur la confiance — à savoir les questions 6 à 11 présentées précédemment. Enfin certains auteurs modifient plus ou moins le questionnaire proposé par Jian. C’est le cas de Chen qui contextualise les questions en nommant le système. Ainsi au lieu de dire “le système est trompeur” (question 1 chez Jian), il “le RoboLeader est trompeur” dans le cadre de ses expérimentations sur le contrôle multi-UGV[22] [17, 18]. Quant à Bowman [10], il adapte le questionnaire dans le cadre des systèmes réseaux. Il reformule les questions afin de distinguer le réseau en tant qu’entité et les services fournis par celui-ci. Nous avons alors des questions telles que “le réseau est digne de confiance” (l’entité) ou “j’étais capable d’envoyer des messages” (le service).

Les données issues de questionnaires sont généralement agrégées à l’aide d’une moyenne afin d’établir la variable de confiance. Comme on a pu le voir dans la partie précédente, certaines questions sont directement liées à la confiance et d’autres à la méfiance. Ces dernières sont intégrées dans le calcul de la moyenne par inversion d’échelle de valeurs : $\text{Valeur}_{inverse} = \text{Valeur}_{max} - \text{Valeur}_{mesure} + 1$ (pour une échelle allant de 1 à Valeur_max).

[21] Cet outil statistique permet de valider les relations des différentes variables du modèles.
[22] Unmanned ground vehicle, véhicule terrestre autonome

2.5.2 Mesure objective

Il n'y a que très peu de mesures objectives de la confiance. Nous pouvons citer une mesure mise au point par Freedy [47] qui est basée sur l'observation des comportements de l'opérateur et du système. En effet, l'idée est d'évaluer par un score de qualité[23] afin de savoir à qui est attribué une tâche. L'évaluation de confiance devient donc le ratio entre le score pour la non-attribution de la tâche au système sur le nombre d'intervention réalisé par l'opérateur. Ainsi, une tâche qui devrait être attribué au système et où l'opérateur n'intervient pas aura un score élevé. Cela soutient nos précédentes critiques de mettre l'interaction au cœur du modèle de la confiance afin de pouvoir l'évaluer. Cette approche est validé expérimentalement en comparant les scores obtenus à une évaluation de la confiance basée un questionnaire — à savoir celui utilisé par Lee et Moray [64].

Le défaut d'une telle mesure est qu'elle ne tient pas compte du contexte et repose uniquement sur la fréquence d'utilisation de l'automate. Un opérateur qui s'ennuie pourrait ainsi intervenir pour s'occuper, et un opérateur en sur-charge devra déléguer une partie de ses tâches à l'automate qu'il ait confiance ou non. La mesure réalisée ici, se base sur une fréquence d'intervention de l'opérateur sans essayer de comprendre pourquoi il intervient. Ainsi, une telle approche qui prendrait en considération le dialogue Homme-machine amenant à cette intervention pourrait pallier à ces défauts.

2.5.3 Nouvelle approche

Au regard du modèle de Lee [66], nous voyons que la confiance influencera la formulation des intentions et par conséquent les actions de l'opérateur. C'est à partir cette même idée que Freedy [47] justifie ses travaux sur une mesure objective de la confiance dans les relations Homme-machine. Il base sa mesure sur des modèles décisionnels de l'allocation de tâche entre l'homme et la machine. Or nous avons insister à plusieurs reprise sur la place que prend l'interaction au sein des modèles de confiance. En effet, nous proposons en ce qui nous concerne une évaluation de la confiance qui repose sur une modélisation des interactions Homme-machine, c'est-à-dire sur un modèle de dialogue.

L'approche du dialogue pour l'évaluation de la confiance permet, en effet, de répondre aux critiques (section 2.5.2) faites précédemment sur la mesure proposée par Freedy [47]. Nous pensons que les motivations de l'opérateur, lorsqu'il intervient, transparaissent au niveau du dialogue. En effet, nous considérons que le dialogue sera minimal si l'opérateur intervient par ennui, et que même en temps de crise l'opérateur aura une stratégie de dialogue qui diffère s'il fait confiance ou non — en prenant, par exemple, le temps de contrôler l'activité du système.

Enfin, l'avantage principal d'une mesure objective est qu'elle se base sur des observables de sorties, c'est-à-dire sur les conséquences de la confiance — lors de l'interaction en ce qui nous concerne. Les questionnaires, quant à eux, reposent

[23] a goodness score

sur des modèles théoriques qui abordent la confiance par ces causes — à savoir les facteurs présentés à la section 2.2.1. Ainsi nous ne souhaitons pas simuler le processus de confiance mais plutôt mesurer ses conséquences et estimer le degré de confiance qui en est à l'origine.

3

Dialogue

Au chapitre précédent, nous avons abordé la confiance. Nous avons ainsi vu, au travers de quelques modèles, que l'interaction y avait une place centrale. Nous pensons que le dialogue Homme-machine peut, par sa forme, être représentatif de la confiance d'un opérateur vis-à-vis de son système. C'est pourquoi nous proposons d'aborder maintenant un certain nombre de modèles de dialogue, afin d'y trouver une approche du dialogue qui nous permette d'analyser cette forme, et ainsi d'évaluer la confiance de l'opérateur.

3.1 Différentes approches du dialogue

Dans la littérature des approches du dialogues [21, 26, 46], nous pouvons distinguer un certain nombre d'approches différentes qui sont regroupées en plusieurs catégories : les grammaires qui reposent sur l'analyse syntagmatique de la langue, la planification qui fait appel aux outils informatiques du même nom, la collaboration qui met en avant les représentations mutuelles des interlocuteurs et l'information qui est centrée sur l'évolution des connaissances partagées par les interlocuteurs.

3.1.1 Approche par la grammaire

L'approche par la grammaire repose sur une analyse syntagmatique de la langue, c'est-à-dire sur une analyse de sa structure syntaxique, sa grammaire. Cette approche commence dans les années 70 avec les travaux de Sinclair et Coulthard [104] qui décrivent le dialogue entre enseignant et élèves au sein d'une classe. Tout en bas de la classification, nous trouvons les actes de langage (poser une question, répondre, informer, *etc*) qui sont structurés les uns par rapport aux autres par les mouvements[1] au niveau supérieur. Ces derniers sont des séquences que le dialogue suit : une question amène à une réponse, une proposition à un acquittement *etc*. Ces séquences sont comparables à des paires adjacentes qui sont issues de l'attente d'une réponse à une question, d'un acquittement à une proposition, *etc*. Elles traduisent donc les contraintes séquentielles et hiérarchiques du dialogue. Jefferson [53] identifie un certain nombre de paires adjacentes qui, ensemble, permettent de modéliser le dialogue, et forment donc sa grammaire.

[1] move en anglais

Si les paires adjacentes définissent la forme du dialogue, les actes de langage en définissent le contenu. Ces derniers sont définis par Austin [5] selon six niveaux. Les quatre premiers concernent la réalisation de l'acte de langage : l'acte phonétique (formulation des phonèmes), l'acte phatique (formulation de mots), l'acte rhétique (utilisation des mots dans une certaine signification) et l'acte locutoire (énonciation quelque chose, indépendamment du sens) qui est l'aboutissement des trois niveaux précédents. A ces niveaux sont ajoutés l'acte illocutoire (le sens que l'on veut transmettre) et l'acte perlocutoire (l'effet voulu sur notre interlocuteur). Ainsi, si nous examinons l'exemple suivant : "Asseyez-vous, je vous prie.", nous avons alors :

- Acte phonétique : former les sons [aseje] (illustré uniquement pour le premier mot) ;
- Acte phatique : énoncer les mots "asseyez", "vous", "je", "prie" ;
- Acte rhétique : employer les mots "asseyez", "vous", "je", "prie" avec une certaine signification ;
- Acte locutoire : dire "Asseyez-vous, je vous prie" ;
- Acte illocutoire : prier son interlocuteur de s'asseoir ;
- Acte perlocutoire : essayer de faire s'asseoir son interlocuteur en lui faisant peur ou en lui faisant se conformer à une norme sociale par exemple.

Searle [100] complète cette théorie en définissant une taxonomie des actes illocutoires. Celle-ci est composée de cinq classes :

- les actes assertifs (ou représentatifs), pour que l'interlocuteur accède à une connaissance e.g. une affirmation ;
- les actes directifs, pour que l'interlocuteur fasse quelque chose par exemple une question, un ordre ;
- les actes promissifs, par lesquels que le locuteur s'engage à une future action par exemple une promesse, une invitation ;
- les actes expressifs, pour exprimer un sentiment vis-à-vis de l'interlocuteur par exemple une excuse, des remerciements ;
- les actes déclaratifs, pour exprimer une sanction ou un changement d'état d'un point de vue institutionnel par exemple déclaration de guerre, nomination.

Il est à noter que les grammaires peuvent être modélisées sous forme de machines à états dans lesquelles ces actes illocutoires sont associés aux transitions. Les états, quant à eux, représentent l'état du dialogue. Par exemple, un état atteint après une question aura pour condition de sortie un acte illocutoire assertif — c'est-à-dire une réponse.

La grammaire définit donc la forme du dialogue et ne permet aucun écart au modèle. Cependant, une flexibilité est apportée par l'utilisation de frames. En effet, ces dernières, définies par Minsky, permettent de décrire des situations ou des objets stéréotypes [75]. Il est ainsi possible de décrire les relations entre les différentes informations qui forment le fond du dialogue, le domaine, à l'exemple de Hulstjin et *al.* [52] qui modélisent un système de réservation pour des pièces de théâtre. Par exemple, dans notre propre domaine d'étude, un frame peut définir, pour le plan de vol d'un drone, une liste de points de passage qui sont à leur tour définis par un

frame. Les règles de grammaires ne décrivent alors que le processus de dialogue tandis que les frames donnent la description du dialogue. Cette distinction est introduite par Aust [4] qui propose en lieu et place des frames un langage de description du dialogue[2]. Cela permet une plus grande souplesse dans l'ordre de traitement de l'information mais n'autorise toujours aucune variabilité dans le comportement de l'opérateur puisque le processus de dialogue est encore défini par des règles.

De notre point de vue, l'approche par la grammaire est très restrictive. En effet, le modèle définit rigoureusement la forme du dialogue et ne permet pas à un opérateur des écarts de dialogue. Ainsi, il n'est pas possible d'observer des variations dans son comportement, et d'établir une évaluation de la confiance exprimée par celui-ci. Pour obtenir ce résultat, il serait en effet utile de pouvoir inférer un modèle de l'opérateur au cours du dialogue.

3.1.2 Approche par la planification

Un aspect absent de l'approche par grammaire est la prise en compte de l'acte perlocutoire. Pour cela, il faudrait inférer les intentions et les objectifs du locuteur. Or, une grammaire seule ne le permet pas. C'est pourquoi la planification a été proposée comme outil de modélisation du dialogue. Cette approche repose sur plusieurs éléments :

- les actions : actes de dialogues ou actions annexes liés au dialogue dont l'exécution affecte les croyances (ce que croit connaître un individu), les objectifs et les intentions des interlocuteurs ;
- les plans d'actions : listes des actions nécessaires à la réalisation des objectifs ;
- les états mentaux : ils représentent les croyances (ce que croit connaître un individu) ainsi que les intentions et les objectifs des participants ;
- les outils de reconnaissance de plan : ils permettent d'identifier les plans de l'interlocuteur à l'aide de règles d'inférence ;
- les outils de planification : ils construisent le plan du locuteur à partir de ses croyances, de ses intentions et des actions de l'interlocuteur, à l'aide de règles d'inférence.

Les outils de reconnaissance de plan permettent au gestionnaire de dialogue d'établir un modèle de l'interlocuteur, à savoir ses intentions et ses objectifs. Les outils de planification, quant à eux, déterminent ensuite les réponses à apporter pour répondre au but de l'interlocuteur [25].

Dans le cadre du contrôle supervisé, une utilisation de cette approche est réalisée dans le projet WITAS[3] qui avait pour objectif de montrer un UAV capable entre autres d'interagir avec un opérateur en langage naturel. Le gestionnaire de dialogue du projet se décompose en plusieurs parties [68] :

- Arbre des mouvements de dialogue[4], qui décrit l'état courant du dialogue et

[2] DDL, Dialogue description language en anglais
[3] Wallenberg Laboratory for Information Technology and Autonomous Systems
[4] *Dialogue move tree*

dont chaque nœud représente un acte de dialogue ;
- Arbre d'activité[5], qui représente le plan d'action de l'UAV ;
- Agenda du système[6], qui représente ce qui doit être communiqué à l'opérateur ;
- Liste d'attentes[7], qui représente les questions du système en attente de réponse de l'opérateur ;
- Liste d'éléments saillants[8], qui permet la résolution des relations anaphoriques et déictiques.

Ces différentes parties permettent d'avoir une représentation du dialogue dont le processus se divise en deux étapes. La première "l'association"[9] consiste à interpréter les actes de dialogues de l'opérateur et à les intégrer à l'arbre des mouvements de dialogue (fig.3.1). La seconde est le traitement des nœuds de l'arbre, à savoir la mise à jour du plan d'action du robot et de l'agenda. Enfin un générateur de dialogue crée les actes de dialogues à réaliser à partir de l'agenda.

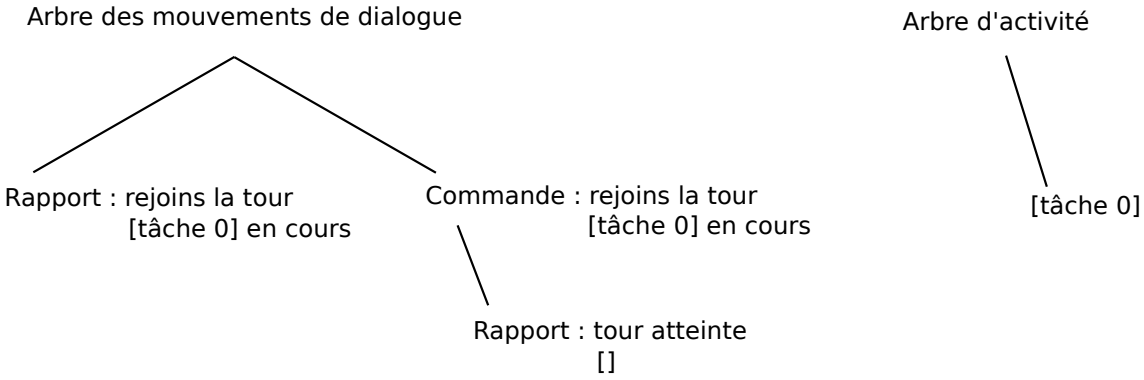

FIG. 3.1 : Exemple d'arbre de mouvements de dialogue et d'arbre d'activité du projet WITAS [68].

On retrouve dans cet exemple les fonctions de reconnaissance de plan, et de planification, dans le processus de mise à jour du plan d'actions du système. En effet, lors de la planification des activités de l'UAV, les intentions et objectifs de l'opérateur sont inférés au plan par l'intermédiaire des commandes qu'il a données.

Les intentions de l'opérateur ainsi que ses objectifs correspondent aux mises à jour du plan d'actions de l'UAV. L'objectif du gestionnaire de dialogue de WITAS n'est pas la prise en compte du facteur humain dans la gestion du dialogue, mais la mise en œuvre d'un langage naturel et multimodal dans une application de contrôle supervisé. C'est pourquoi la représentation des états mentaux au-delà des objectifs (à savoir les commandes de l'opérateur) n'est pas prise en compte.

[5] *Activity tree*
[6] *System Agenda*
[7] *Pending list*
[8] *Salience list*
[9] *attachment*

L'approche par la planification attaque le problème de représentation de l'opérateur, mais dans un objectif de tâche, comme nous le constatons dans le projet WITAS. En effet, la reconnaissance de plan est utilisée ici pour comprendre l'acte perlocutoire en termes de tâches et/ou d'actions attendues par celui-ci. Ainsi, bien que cette approche s'intéresse aux intentions, elle n'aborde pas une représentation complète des états mentaux de l'opérateur. De ce fait, il faudrait aborder un modèle de dialogue où les représentations mutuelles auraient une place plus importante. En effet, il ne faut pas oublier qu'ici le dialogue est considéré comme une activité conjointe, c'est-à-dire quelque chose que les participants font ensemble [27].

3.1.3 Approche collaborative

Nous avons pour le moment abordé deux approches. La première, les grammaires, modélisent la forme du dialogue et sont, de notre point de vue, trop restrictives pour permettre d'y observer des variations de comportement de la part de l'opérateur. Quant à l'approche par la planification, bien qu'elle soit censée incorporer une représentation mutuelle des interlocuteurs, elle n'établit, dans sa forme courante, qu'une représentation des intentions et des objectifs des interlocuteurs. Cette approche nous semble intéressante mais insuffisante. En effet, pour évaluer la confiance, les intentions seules de l'opérateur ne suffisent pas. Dans le cadre du contrôle supervisé, l'opérateur peut formuler les mêmes intentions d'usage vis-à-vis d'un automate sans pour autant avoir un même niveau de confiance vis-à-vis de celui-ci. C'est pourquoi nous pensons que la formulation de ces intentions prendra des formes différentes selon sa confiance. Pour observer cela, l'approche collaborative du dialogue s'axe davantage sur une représentation mutuelle des participants et non sur une tâche à réaliser — même si *in fine* l'objectif est la réalisation de cette tâche.

La notion de dialogue comme activité conjointe est définie par Novick *et al.* [83] dans un modèle de collaboration pour l'élaboration d'informations mutuelles spécifiques à une tâche. Les agents (individus ou systèmes) participent au dialogue et construisent un ensemble de connaissances partagées à l'aide d'actes de dialogues indépendants et spécifiques à un domaine. Un agent A établit, à partir de ses connaissances et de ce qu'il croit qu'un agent B connaît, une intention. Celle-ci est réalisée par une action qui est généralement la formulation d'un énoncé. L'action est alors interprétée comme un acte spécifique par l'agent B, en accord avec ses connaissances et ce qu'il croit que l'agent A connaît. Bien que les connaissances des deux agents semblent séparées l'une de l'autre, Novick propose un niveau de connaissances mutuelles, i.e. connaissances partagées. On retrouve le même concept chez Traum [108] qui définit le modèle d'agent conversationnel en étendant l'architecture BDI[10] d'un agent par l'incorporation du concept de connaissances mutuelles, i.e. ce que les deux agents dialoguants croient tous deux comme vrai. Ainsi les agents décident de leurs actions en fonction de leur propre but mais aussi des connaissances qu'ils

[10]Le modèle Belief Desire Intention décrit comment les actions d'un agent dans le monde affectent ses connaissances.

ont de l'autre. Nous retrouvons ici, le même processus que chez Lee (fig.3.2) où les connaissances pilotent les intentions et les actions de l'opérateur. A ceci près qu'entre les connaissances et les intentions, il y a l'évolution de la confiance. Cela signifie que les modèles collaboratifs à base de connaissances partagées ouvrent une fenêtre intéressante sur la confiance des interlocuteurs.

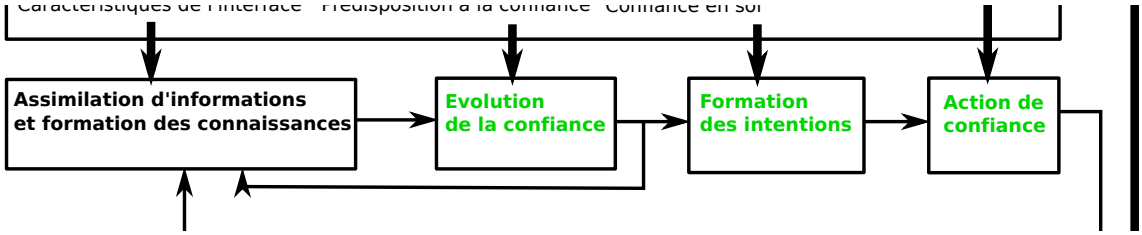

FIG. 3.2 : Extrait du modèle descriptif de la confiance par Lee avec en vert les intentions influencées par la confiance qui mènent aux actions de l'opérateur [66].

L'approche collaborative permet de modéliser les échanges entre agents qui, par le dialogue, établissent des coopérations afin de réaliser leurs objectifs respectifs. Les propositions que se font les agents sont généralement acceptées ou rejetées de manière binaire. Chu-Carroll [20] va plus loin en abordant au sein du dialogue les stratégies de négociation. Afin de décider d'un accord ou non concernant une proposition faite par un interlocuteur, son modèle construit une représentation des connaissances de l'interlocuteur afin d'argumenter et de convaincre avec une contre-proposition. La possibilité d'une négociation introduit une variabilité dans le comportement des agents selon la manière dont ils négocient. Dans ce modèle, l'objectif est de trouver un terrain d'entente pour une coopération entre agents, mais nous pouvons faire le parallèle avec un opérateur qui agit de différentes manières avec son système. Ainsi l'opérateur modifie son comportement selon les connaissances qu'il a du système, ce qui concorde avec les modèles de la confiance où les connaissances sur les compétences et la prédictibilité *etc.* influencent la confiance, et par incidence les intentions de l'opérateur.

L'évaluation de la confiance basée sur l'analyse du dialogue nécessite une représentation mutuelle des interlocuteurs. En effet, dans le cadre du contrôle supervisé, ce sont les perceptions que l'opérateur a du système qui vont déterminer son degré de confiance. Ces perceptions s'établissent au cours de l'interaction Homme-machine et donc au travers des informations partagées par le système avec l'opérateur. Il est vrai que, selon le modèle de Lee (fig.3.2), la confiance est déterminée par les connaissances qu'a l'opérateur du système ainsi que par un certain nombre de paramètres environnementaux. Il serait alors peut-être préférable non pas d'inférer un modèle de l'opérateur pour déterminer sa confiance mais d'analyser la manière dont l'opérateur s'approprie les informations du système. En effet, il paraît rationnel que le degré de confiance agisse sur le volume d'informations qu'essayera d'obtenir l'opérateur. Autrement dit, un opérateur peu confiant ou en début de relation cherchera sûrement

à acquérir une grande quantité d'informations afin de décider d'une intervention ou non. Il faut donc nous intéresser à un modèle de dialogue centré sur le partage de l'information et sur les connaissances partagées.

3.1.4 Approche du dialogue basée sur l'information

L'approche basée sur l'information, définie par Traum, ne cherche qu'à décrire l'évolution de l'état des connaissances partagées au cours d'un dialogue. De plus, cette approche dissocie complètement le dialogue de la tâche. Ainsi l'état des connaissances partagées représente les informations échangées au cours du dialogue. Il est spécifique à chaque dialogue car il contient l'accumulation des précédentes actions du dialogue. Il motive aussi les actions futures. Par exemple, une phrase ajoute généralement une information propositionnelle, une question motive l'interlocuteur pour énoncer une phrase. Généralement, mais non nécessairement, l'état de l'information des participants d'un dialogue représente les informations qu'ont les participants à un moment du dialogue — celles qu'ils ont apportées et celles qu'ils ont retenues — et leurs motivations pour agir dans le futur.

Un gestionnaire de dialogue basé sur l'information consiste pour Traum en :

- une description des composants informationnels, incluant les aspects du contexte aussi bien que les facteurs de motivation (e.g., les participants, le champ commun, les structures linguistiques et intentionnelles, les obligations et les engagements, les croyances, les intentions, etc.) ;
- des représentations formelles des composants précédemment cités (e.g. tels que des listes, des ensembles, des structures, des enregistrements, etc.) ;
- un ensemble d'actes contributifs au dialogue, qui déclenchent les mises à jour de l'état de l'information. Ceux-ci sont généralement corrélés à des actions externes au gestionnaire, telles que des énoncés en langage naturel. Il est donc parfois nécessaire d'avoir des outils de reconnaissance pour ces actes contributifs ;
- un ensemble de règles de mise à jour qui gouverne la mise à jour de l'état de l'information, selon l'état courant de l'information et l'acte contributif réalisé. Cet ensemble inclut aussi les règles de génération d'actes contributifs (lorsque le système participe au dialogue et ne se contente pas de le gérer) ;
- une stratégie de mise à jour pour décider quelle règle est à appliquer à un moment donné. La stratégie peut être des plus simples (e.g. "prendre la première règle") ou faire appel à des mécanismes plus complexes comme la théorie des jeux, la planification, ou des méthodes statistiques.

En faisant de l'information la base du modèle, cette approche permet une grande souplesse dans la gestion du dialogue, puisqu'il est possible de créer l'équivalent d'une grammaire ou d'une approche par la planification, à l'aide des règles de mises à jour et de leur stratégie d'application. Elle offre donc une grande variabilité dans les interactions permises à un opérateur, dans la limite des règles de mises à jour qui auront été définies. Or, nous pensons que les choix de l'opérateur pour interagir avec

le système seront liés à sa confiance. Et donc comme nous l'avons dit, à la section 3.1.3, l'évaluation de la confiance devrait reposer sur la construction et l'évolution d'un champ commun entre l'homme et la machine. Intéressons-nous donc à ce modèle de plus près, et plus particulièrement la théorie sous-jacente qui décrit le processus de construction et de maintenance du champ commun.

3.2 Théorie du grounding

Les dernières approches de la section précédente font référence à une notion importante, qui attire notre attention : les connaissances partagées (le champ commun). En effet, l'élément central d'un dialogue, comme le propose le dernier modèle, est le partage d'informations. Or, dans notre problématique sur l'évaluation de la confiance, c'est à partir du dialogue avec le système que la confiance d'un opérateur évoluera. C'est donc cette notion que nous abordons à présent.

3.2.1 Champ commun

Clark définit la notion de champ commun comme la somme des connaissances — ainsi que croyances et hypothèses — communes, mutuelles et conjointes de plusieurs individus. Il en propose différentes représentations [22] :

- représentation basée sur le partage : une information p fait partie du champ commun d'une communauté C si :
  1. tous les membres de C possèdent un ensemble de connaissances de base b.
  2. b indique à chaque membre de C que tous les membres de C connaissent b.
  3. b informe la communauté C de l'information p.

  L'initialisation d'une base pour le champ commun est alors explicite et se rattache à la perception du contexte par exemple.
- représentation réflective : issue de la représentation basée sur le partage. Cette dernière est modifiée en supprimant la référence à un ensemble de connaissances de base, c'est-à-dire en supprimant la première condition. L'information p fait alors partie du champ commun à la condition suivante :
  1. les membres de la communauté C connaissent à la fois p et la condition '1'.

  Cette représentation est réflective car elle fait référence à elle-même. En effet, une information intègre le champ commun à partir du moment où chaque membre d'une communauté a connaissance de l'information et qu'ils ont connaissance du fait que chacun d'eux la connaît.
- représentation itérative : une information p fait partie du champ commun si :
  1. les membres de C possèdent l'information p.

2. les membres de C possèdent l'information que les membres de C possèdent l'information p
3. les membres de C possèdent l'information que les membres de C possèdent l'information que les membres de C possèdent l'information p.
4. les membres de C possèdent l'information que les membres de C possèdent l'information que les membres de C possèdent l'information que les membres de C possèdent l'information p.
5. et ainsi de suite...

La représentation utilisée par tout un chacun est la première, et c'est aussi celle qui sera employée pour les modèles informatiques du dialogue. Mais pour cela, il faut une base initiale de connaissances qui est issue du contexte (comme dit précédemment), mais qui se construit aussi à partir de notre culture (normes sociales), de notre expertise (éducation, formation) et des évidences d'une appartenance à une même communauté (deux personnes travaillant à un même poste sauront qu'elles partagent une expertise technique).

Enfin, le champ commun est essentiel à la coordination d'actions conjointes. En effet, c'est à l'aide d'un accord explicite (faisant donc appel à la représentation basée sur le partage) que les actions se coordonnent. Ces accords sont donc intégrés au champ commun lorsqu'ils sont établis.

3.2.2 Elaboration du champ commun

Selon Clark [22], le dialogue est une activité conjointe dont la réussite est certifiée par une suffisamment bonne intégration à un champ commun des informations échangées au cours du dialogue. Par ailleurs, le principe de clôture de Norman (1988) dit que des agents réalisant une action requièrent une preuve, suffisante pour leur objectif, de la réussite de l'action. En d'autres mots, pour qu'un dialogue réussisse, il faut que le locuteur perçoive de manière évidente que l'activité conjointe qu'il a initiée a réussi. Dans le cadre de l'activité conjointe, ce principe est mutuel : le locuteur comme l'interlocuteur doivent percevoir l'évidence de la clôture de leur activité conjointe.

Ce principe pose les bases du modèle de contribution introduit par Clark et Schaefer [23]. Ce modèle décrit l'activité conjointe qu'est le dialogue en deux phases :

- la phase de présentation, au cours de laquelle une information est introduite et soumise à l'interlocuteur,
- la phase d'acceptation, où l'interlocuteur acquiesce à la réception de l'information.

Ce modèle peut se représenter, selon Traum [108], sous forme d'un réseau de transitions (fig.3.3) dont l'état final implique l'apport d'un nouvel élément (information) au champ commun. A partir de l'état initial ('S'), on passe à un état intermédiaire ('1') lorsqu'un locuteur effectue une énonciation apportant une nouvelle information. Enfin, lorsque son interlocuteur énonce son acceptation, le réseau de transition

bascule dans un état final ('F').

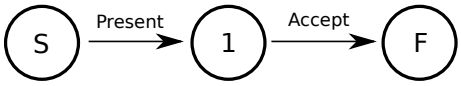

FIG. 3.3 : Réseau de transitions pour le contribution model de Clark and Schaefer [108]

Cette dernière phase peut prendre plusieurs formes qui, selon les travaux de Roque et Traum [96], induisent différents degrés d'appartenance de l'information au champ commun (le degré de grounding). La forme que prend l'acceptation va traduire un degré d'appartenance de l'information au champ commun plus ou moins important. Roque distingue neuf degrés de grounding :

1. inconnu : l'information n'a pas encore été présentée,
2. mal compris : l'introduction de l'information enclenche un processus de clarification,
3. non acquitté : l'information a été présentée et aucune réponse n'est formulée,
4. accessible : l'information vient d'être présentée,
5. acquittement du signal : l'information a été présentée et un acquittement a eu lieu en retour,
6. acquittement renforcé du signal : l'information a été présentée, un acquittement a eu lieu en retour et le dialogue se poursuit toujours sur la même information,
7. acquittement du contenu : l'information a été présentée et un acquittement a eu lieu par une reformulation du contenu,
8. acquittement renforcé du contenu : l'information a été présentée, un acquittement a lieu par reformulation du contenu et le dialogue se poursuit toujours sur la même information.
9. assumé : l'information a intégré le champ commun par une autre méthode.

Le degré de grounding est une notion intéressante au sens où, lors d'une conversation, un locuteur exigera un degré minimum pour le grounding d'une information, selon l'importance qu'il y attache. Nous retrouvons l'idée de "suffisance" du principe de clôture[11]. Le mécanisme de grounding ne garantit pas une compréhension commune de l'information mais juste un partage de cette dernière [19]. En effet, une information, dont le degré de grounding est l'acquittement du signal, peut ne pas avoir été comprise correctement. Dans ce cas, seule la réception du signal est confirmée.

[11] Des agents réalisant une action requièrent une évidence, suffisante pour le but courant, qu'ils ont réussi leur action.

3.2.3 Modèle récursif de la théorie du grounding

Traum part du modèle de contribution afin de développer un modèle informatique de la théorie du grounding. Pour cela, il généralise le modèle de manière à pouvoir modéliser n'importe quel dialogue. Ce dernier se divise en unités de discours qui représentent une séquence de dialogue initiée par l'ajout/présentation d'une nouvelle information au champ commun. Le locuteur qui énonce la nouvelle information est l'initiateur (I) et son interlocuteur qui lui répond est le répondant (R).

Pour le formalisme des représentations graphiques, la lettre (I ou R) représente l'agent déclencheur de la transition.

3.2.3.1 Enonciation continue

La première modification apportée par Traum permet de décrire la conversation suivante :

I : Move the box car to Corning
I : and load it with oranges
R : ok

Tandis que pour Clark et Schaefer, le dialogue aurait nécessité une validation de chaque énoncé :

I : Move the boxcar to Corning
R : ok
I : and load it with oranges
R : ok

Une transition "continue" permet au locuteur initial d'ajouter de nouveaux énoncés à son propos en bouclant sur l'état '1' du réseau de transitions. L'ensemble des énoncés sera intégré par une unique acceptation (fig.3.4).

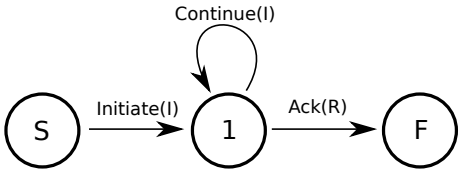

FIG. 3.4 : Réseau de transitions initial d'une unité de discours [108]

3.2.3.2 Retrait d'énoncé

Le locuteur initial peut vouloir se rétracter après avoir fini son énoncé. C'est pourquoi Traum ajoute un nouvel état à son réseau : "death state" ('D'). De plus, les transitions "cancel" amenant à cet état peuvent aussi traduire l'impossibilité de mise en partage de l'énoncé dans le champ commun, en raison de l'évolution de la conversation (fig.3.5).

3.2.3.3 Correction de l'énoncé

Le locuteur possède déjà la possibilité de compléter l'énoncé initial : transition "continue". Mais il peut aussi vouloir modifier et/ou corriger cet énoncé. C'est ce que représente la transition "repair" (fig.3.5).

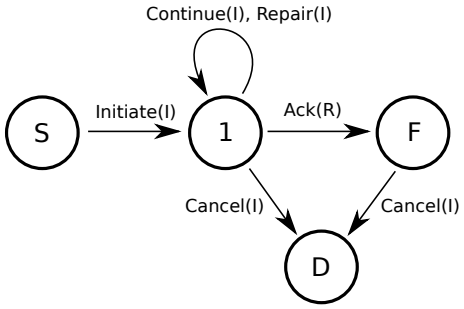

FIG. 3.5 : Réseau de transitions d'une unité de discours avec retrait et autocorrection de l'énoncé [108]

Ces corrections ne dépendent pas uniquement de l'initiative du locuteur initial, mais elles peuvent aussi émaner de son interlocuteur, qui peut effectuer lui-même une correction ou en requérir une. Pour ce faire, Traum ajoute au réseau des transitions récursives (en lettres capitales sur les représentations graphiques, fig.3.6). Les transitions résultantes "REPAIR" et "REQ-REPAIR" reviennent à parcourir un sous-réseau de transitions (respectivement fig.3.7 et fig.3.8), de son état initial (respectivement 'RS' et 'RRS') à son état final (respectivement 'RF' et 'RRF'). Si ce sous-réseau aboutit à l'état "death" (respectivement 'RD' et 'RRD'), alors la transition REPAIR ou REQ-REPAIR est considérée comme n'ayant jamais eu lieu.

Il faut également noter qu'au sein de ces sous-réseaux, les rôles d'initiateur et de répondant sont redistribués.

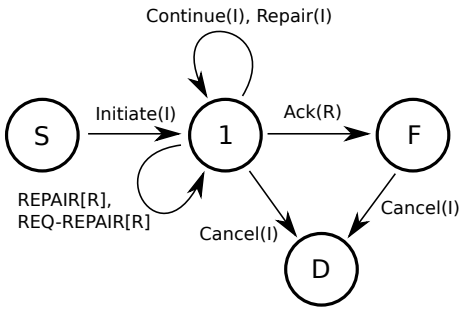

FIG. 3.6 : Réseau de transitions d'une unité de discours avec corrections récursives [108]

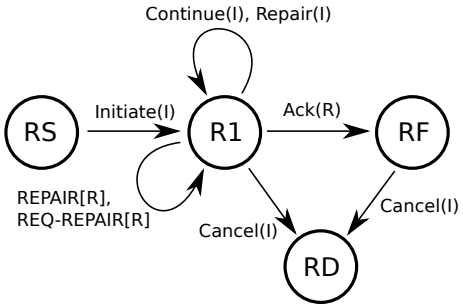

FIG. 3.7 : Sous-réseau de transitions d'une transition REPAIR [108].

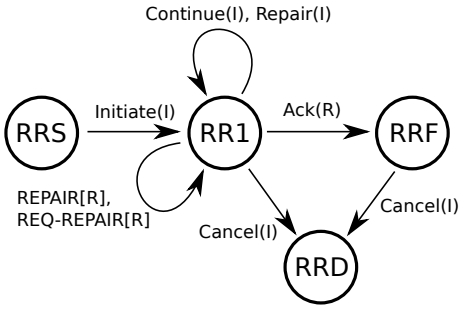

FIG. 3.8 : Sous-réseau de transitions d'une transition REQ-REPAIR [108]

3.2.3.4 Réouverture d'une unité de discours

Enfin, la dernière fonctionnalité ajoutée par Traum permet de rouvrir une unité de discours à partir de son état final. En effet, Traum considère que pendant un court moment une transition corrective peut être réalisée, faisant quitter l'état final du réseau pour son état '1' (fig.3.9).

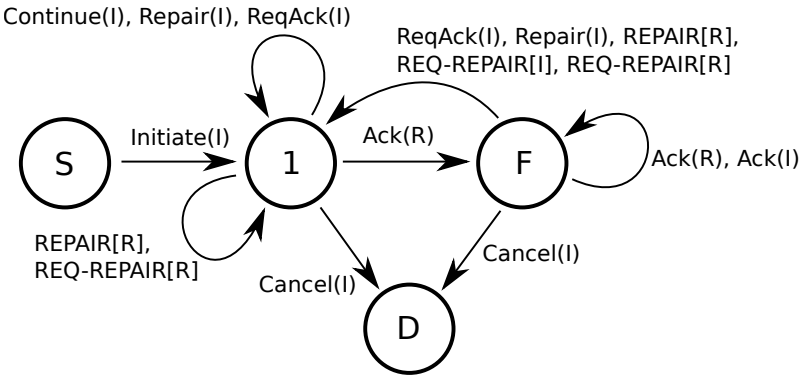

FIG. 3.9 : Réseau de transitions modifié avec corrections récursives [108]

3.2.4 Modèle non récursif de la théorie du grounding

Pour finir, Traum réalise quelques transformations sur son modèle, pour casser les récurrences et en faire un modèle à états finis.

L'état intermédiaire '1' est alors interprété par Traum comme représentant l'initiateur qui a pris l'initiative du discours et ce sans obligation pour l'initiateur de faire évoluer le dialogue. Traum introduit de nouveaux états : l'état '2' représente l'initiateur ayant l'initiative avec l'obligation d'apporter une correction au dialogue ; l'état '3' attribue quant à lui la prise d'initiative au répondant sans aucune obligation pour l'évolution du dialogue ; enfin l'état '4' attribue l'initiative au répondant avec l'obligation d'apporter une correction au dialogue.

Pour finir, un certain nombre de transitions sont adaptées en conséquence (fig.3.10 à fig.3.14). Le tableau (tab.3.1) synthétise l'ensemble des transitions du modèle de grounding tel que défini par Traum [108].

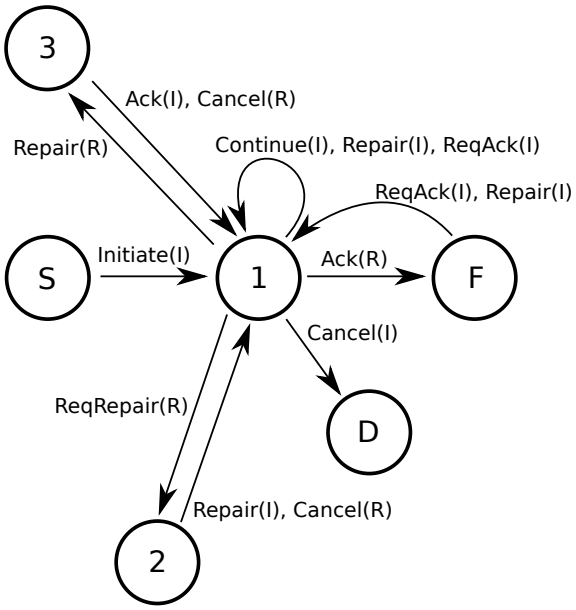

FIG. 3.10 : Réseau de transitions "à partir de" et "vers" l'état 1 [108]

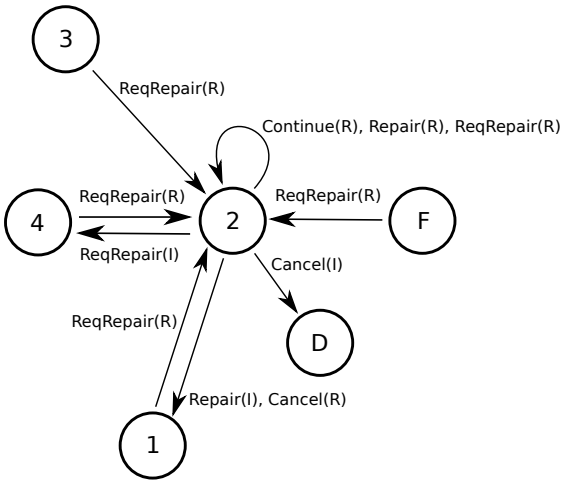

FIG. 3.11 : Réseau de transitions "à partir de" et "vers" l'état 2 [108]

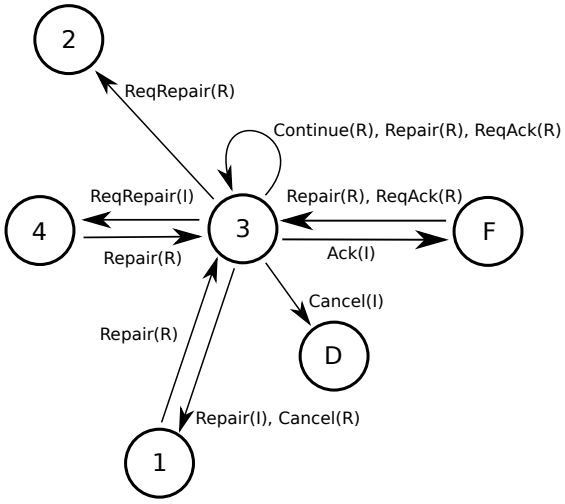

FIG. 3.12 : Réseau de transitions "à partir de" et "vers" l'état 3 [108]

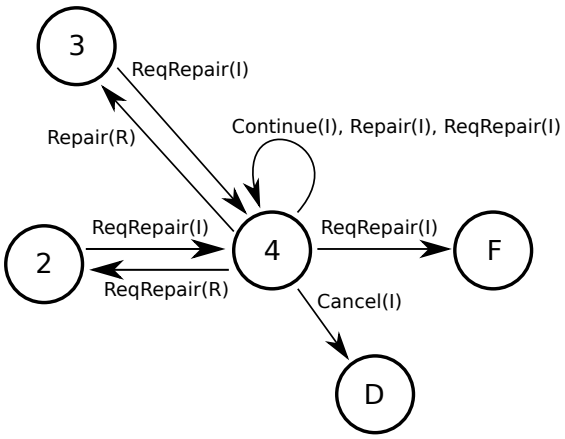

FIG. 3.13 : Réseau de transitions "à partir de" et "vers" l'état 4 [108]

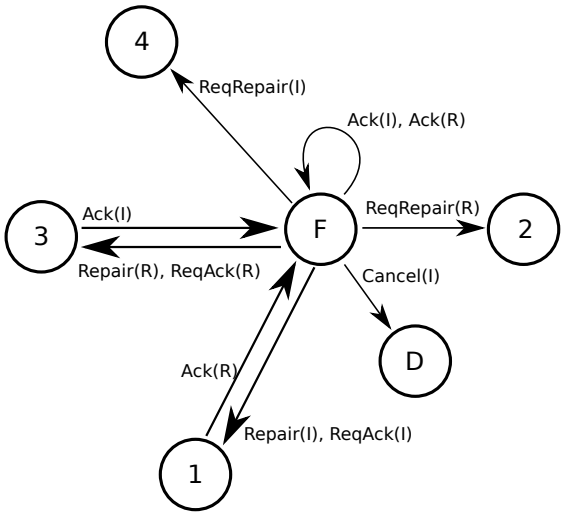

FIG. 3.14 : Réseau de transitions "à partir de" et "vers" l'état F [108]

Next Act	In State						
	S	1	2	3	4	F	D
Initiate	1						
Continue(I)		1			4		
Continue(R)			2	3			
Repair(I)		1	1	1	4	1	
Repair(R)		3	2	3	3	3	
ReqRepair(I)			4	4	4	4	
ReqRepair(R)		2	2	2	2	2	
Ack(I)				F	1*	F	
Ack(R)		F	F*			F	
ReqAck(I)		1				1	
ReqAck(R)				3		3	
Cancel(I)		D	D	D	D	D	
Cancel(R)			1	1		D	

TAB. 3.1 : Table des transitions d'une unité de dialogue [108]. *La demande de correction "reqRepair" est ignorée.

3.3 CONCLUSION

Nous avons présenté un certain nombre de modèles de dialogue afin de voir comment la confiance d'un opérateur pouvait y être perçue. Or, cette analyse part du principe que l'opérateur agit de façon différente selon son degré de confiance. Au regard de l'impact de la confiance au sein du contrôle supervisé, cette hypothèse est tout à fait rationnelle. Cela nous a conduit, en nous basant sur le modèle descriptif de la confiance de Lee, à prendre en considération les modèles de dialogue qui mettent le partage de l'information en avant. Au cœur de cette approche, la théorie du grounding et le modèle qu'en fait Traum ont été retenus.

4 Contrôle multi-drones : expérience préliminaire

Dans les chapitres précédents, nous avons abordé la confiance et son évaluation dans un cadre général. Afin de réaliser nos expérimentations, nous réduisons notre champ applicatif au contrôle multiple d'engins autonomes — et au contrôle multi-drones pour les expérimentations. A travers cette première expérimentation, nous éprouverons notre approche et tenterons d'identifier une relation entre interaction et confiance. De plus, nous pourrons aussi vérifier l'hypothèse formulée au cours de la définition de l'objet de confiance — à savoir que la confiance ne s'adresse pas à un système dans son ensemble mais à chacune de ses fonctionnalités.

4.1 Contrôle multiple d'engins autonomes

Dans l'état de l'art sur le contrôle supervisé, nous avons abordé un certain nombre d'applications telles que le contrôle de processus industriels [57], le pilotage d'avions [76, 112], ou le contrôle d'engins autonomes [34, 79]. Pour étudier expérimentalement notre nouvelle approche de l'évaluation de la confiance, nous nous focaliserons sur le contrôle d'engins autonomes.

4.1.1 Domaine d'application

Depuis quelques années, l'emploi d'engins autonomes (unhabited vehicle, UV) s'est fortement accru. Le domaine militaire est actuellement le principal secteur de déploiement de ces engins très divers : aérien (unmanned aerial vehicle, UAV), terrestre (unmanned ground vehicle, UGV), ou maritime (unmanned surface vehicle, USV ; unmanned undersea vehicle, UUV). Les conflits récents montrent parfaitement la place croissante qu'occupent les drones dans le théâtre des opérations militaires et de la surveillance.

Les succès militaires poussent le milieu civil à exploiter lui aussi les capacités de ces engins. C'est pourquoi de nombreuses recherches sont conduites pour permettre l'intégration des UAVs dans les espaces aériens nationaux. Concernant les véhicules terrestres, il nous suffit de prendre par exemple l'emploi de certains robots lors de la catastrophe de Fukushima (2011) pour l'exploration et l'intervention en zone contaminée [2].

4.1.2 Problématique

Comme nous l'avons vu au chapitre 1, l'automatisation des systèmes peut provoquer une dégradation de la conscience de la situation chez l'opérateur. Ce premier problème est encore plus conséquent dans le contrôle multi-UV où la démultiplication des engins autonomes nécessite une plus grande autonomie de ces derniers. Cummings [33] et Squire [106] constatent expérimentalement cette dégradation qui a alors un effet sur les décisions de l'opérateur ce qui peut aboutir à un mauvais usage de l'automate voire à un échec de la mission.

Le deuxième problème concerne la charge de travail. Ce problème est mis en avant par Cummings et Dixon [41] qui observent qu'une charge de travail trop importante dégrade les performances du couple Homme-machine. Or, dans le cadre d'opérations multi-UV, la multiplication des tâches due à l'augmentation du nombre d'engins autonomes entraîne une augmentation conséquente de la charge de travail de l'opérateur. Celle-ci peut amener à un mauvais usage de l'automate : un usage abusif ou insuffisant, les résultats étant très variable selon les cas d'après Parasuraman [85].

Nous abordons pour notre part un autre facteur lié à l'usage des automates : la confiance. En effet, ce facteur est important dans le contrôle supervisé qu'il prend une part importante dans les décisions de l'opérateur et ses stratégies. Or, nous pensons qu'il est nécessaire d'évaluer ce facteur afin d'anticiper des situations à risque. Nous proposons pour cela une nouvelle approche, basée sur le dialogue, de l'évaluation de la confiance. L'objectif de l'expérimentation présentée ci-après est de mettre en avant l'existence d'un lien entre dialogue et confiance pour soutenir notre démarche.

4.2 Plateforme expérimentale

Afin de réaliser une campagne expérimentale préliminaire, nous avons développé un simulateur de contrôle multi-drones pour des missions de surveillance et d'interception. Le simulateur est divisé en trois parties : l'interface utilisateur, le simulateur, les algorithmes pour les fonctions automatisées.

4.2.1 Présentation globale

Le simulateur permet de simuler des missions de surveillance et d'interception. Un opérateur dispose d'une interface sur laquelle il peut visualiser la zone qu'il doit protéger, ainsi que la localisation des UAVs dont il dispose pour cette mission. De plus, des alarmes affichent la localisation des intrus sur l'interface au fur et à mesure que l'opérateur les détecte. Enfin, l'opérateur peut aussi accéder via l'interface à des informations d'état concernant ses UAVs (e.g. niveau de fuel).

Afin de superviser les UAVs, l'opérateur a quatre tâches à réaliser :

– Agrégation : grouper les alarmes qui sont supposées avoir été générées par les détections successives d'un même intrus ;

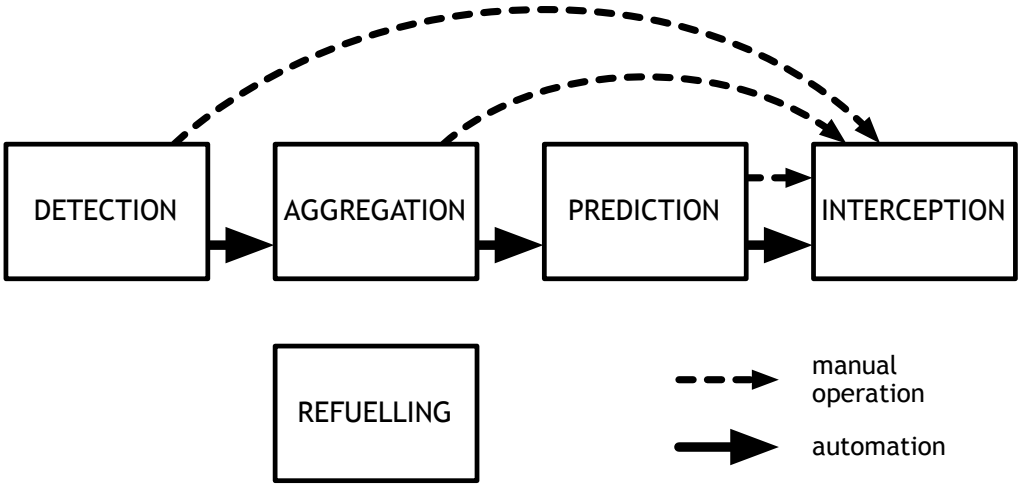

FIG. 4.1 : Relation entre les tâches.

- Prédiction : prédire la trajectoire d'un intrus (en utilisant la position des agrégations d'alarmes) ;
- Interception : sélectionner des UAVs et les positionner de façon à pouvoir intercepter un intrus ;
- Ravitaillement : envoyer des UAVs à la base de ravitaillement lorsque leur niveau de fuel est trop bas.

Chacune de ces tâches peut être gérée par un automate — que l'opérateur peut activer (mode automatique) ou non (mode manuel) — mais un certain nombre d'interdépendances existent entre ces tâches : l'interception automatique utilise l'information de prédiction afin de déterminer la position des UAVs, tandis que la prédiction automatique est basée sur les agrégations d'alarmes (voir fig.4.1).

En mode manuel, l'opérateur peut réaliser l'interception directement avec les informations des agrégations d'alarmes ou les informations des alarmes. Cela signifie qu'il ne réalise pas les commandes manuelles d'agrégation et/ou de prédiction pour réaliser l'interception. Dans ce cas, nous considérons que l'agrégation et la prédiction ont été réalisées mentalement par l'opérateur. A noter que dans ce dernier cas, l'automate ne peut alors prendre aucune initiative pour intercepter un intrus si l'opérateur a désactivé les fonctions automatiques d'agrégation et de prédiction (du fait de leur dépendance).

4.2.2 L'interface

L'interface est divisée en trois parties (fig.4.2) :

le panel tactique : affiche la carte locale avec la position des UAVs et des alarmes ;

le panel informatif : affiche les informations qui concernent les objets sélectionnés (UAV ou alarme) ;

le panel d'action : affiche la liste des différentes commandes et leurs configurations respectives (automatique ou manuelle). Aux quatre tâches est associée une commande correspondante (e.g. "intercept" pour l'interception). Et s'ajoutent à cela deux autres commandes : "go" qui permet d'envoyer un UAV à un point donné de la carte, et "patrol" qui permet de remettre en patrouille un UAV (qui ralliait la base pour un ravitaillement, ou qui était en interception).

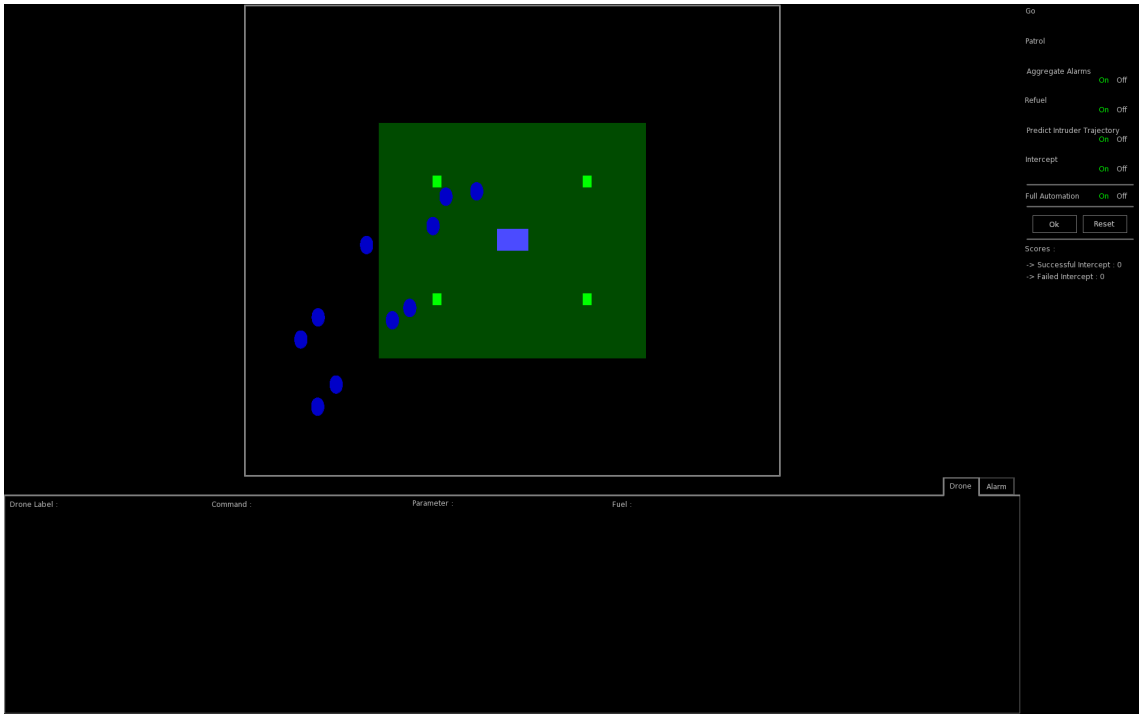

FIG. 4.2 : Interface de l'opérateur.

4.2.3 Le simulateur

Le simulateur a pour rôle de mettre à jour l'état de l'environnement simulé en temps-réel. Cela signifie :

- qu'il calcule les positions des UAVs en fonction des paramètres de leurs commandes (type de commande : Go, Refuel, etc...) et de leur vitesse — que nous avons supposée constante ;
- qu'il met à jour la position des intrus en fonction de la trajectoire planifiée (liste de points de passage définis comme une localisation et un temps) et de leur vitesse (paramètre constant) ;
- qu'il calcule les alarmes de détection à partir de probabilités. Un intrus a une espérance de non détection fixée à une minute au sein du réseau terrestre de

capteurs situé autour de la base à protéger. Ensuite il est indétectable sauf si un UAV le survole, alors l'espérance de non détection étant alors de deux secondes (fig.4.3) ;

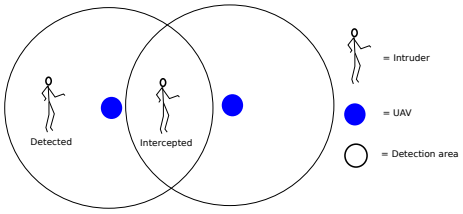

FIG. 4.3 : Détection et interception d'un intrus par deux UAVs.

– qu'il détermine la réussite des interceptions, définie comme la détection de l'intrus simultanément par deux UAVs en mode interception (fig.4.3), c'est-à-dire une double détection. En effet, les deux UAVs forment une barrière en aval de la trajectoire de l'intrus, là où ce dernier est détectable par les UAVs, la double détection permettant alors la confirmation de la présence de l'intrus.

4.3 L'EXPÉRIMENTATION

4.3.1 Panel expérimental

Les sujets forment un groupe de neuf personnes, toutes issues de Télécom Bretagne. Ce groupe est composé de six étudiants, de deux membres du personnel administratif et d'un enseignant-chercheur. Cet échantillon est trop faible pour une étude statistique. Les résultats exposés par la suite ne seront que qualitatifs.

4.3.2 Protocole

Premièrement, l'expérimentation est présentée au sujet comme une étude sur les performances liées à l'interaction. Nous expliquons au sujet que nous voulons mesurer et analyser l'interaction afin de développer de nouvelles méthodes d'interaction pour l'optimisation des performances dans la coopération Homme-machine.

Deuxièmement, le sujet est formé sur le simulateur en deux temps. D'abord, une présentation du simulateur et de son interface est réalisée. Ensuite, le sujet effectue un scénario d'entraînement d'une durée de 15 minutes au cours duquel celui-ci peut être aidé et poser n'importe quelle question sur l'utilisation du simulateur.

Troisièmement, le sujet réalise deux scénarios (décrits en section 4.3.3) durant lesquels il agit sans aide et sans intervention extérieure. L'interaction est mesurée durant cette étape (les mesures seront présentées en section 4.3.4).

Chaque participant réalise les deux mêmes scénarios. Afin de réduire les effets d'apprentissage, nous avons divisé les sujets en deux groupes : le premier groupe commençant avec le scénario n°1, le second avec le scénario n°2.

A la fin des simulations, nous interrogeons le sujet sur sa perception et sur sa stratégie d'utilisation de chaque fonction automatisée. Ainsi nous pouvons évaluer sa confiance sans aborder explicitement cette notion. En effet, nous ne voulons pas biaiser l'expérimentation par l'influence que pourrait avoir une connaissance a priori du but expérimental. C'est pourquoi, et en accord avec les modèles de la confiance vus précédemment, l'entretien est orienté vers les facteurs de la confiance. Ainsi, des questions sont posées sur les performances (compétences perçues) et la prédictibilité de chacun des automates. Enfin, nous demandons aux sujets d'expliquer et de justifier leur comportement vis-à-vis du système. Cela permet de mettre en relation les performances et les prédictibilités perçues avec le comportement effectif du sujet.

4.3.3 Scénarios

Deux scénarios sont joués par les sujets ; ils se différencient par une difficulté spécifique à chacun.

4.3.3.1 Scénario n°1

Le premier scénario a été conçu pour surcharger l'opérateur avec un grand nombre d'intrusions simultanées. Pendant les 15 minutes que dure le scénario, il y a deux groupes de six intrus qui apparaissent (le premier groupe au début du scénario, le second à la moitié).

Pour ce scénario les automates ont une bonne efficacité avec dix UAVs à disposition (75% des intrus sont interceptés automatiquement sans opérateur). Nous avons décidé de débuter ce scénario avec les automates inactifs par défaut.

4.3.3.2 Scénario n°2

Le second scénario est conçu pour mettre en défaut l'automate. En effet, l'automate ne peut pas distinguer le nombre d'intrus en un même point (une surface unitaire dépendante de la résolution utilisée par les algorithmes, c'est-à-dire une case de taille minimale). Ainsi, si l'automate intercepte l'un des intrus, les autres intrus continuent d'avancer jusqu'à ce qu'il y ait suffisamment d'alarmes (deux) pour une nouvelle interception. L'espérance du temps au cours duquel un intrus n'est pas détecté laisse suffisamment de temps à cet intrus pour atteindre sa cible.

Le scénario est divisé en deux parties. La première consiste en l'arrivée de deux groupes de deux individus qui viennent de directions opposées. Les intrus se séparent après quelques minutes pour rejoindre leurs cibles respectives. La deuxième partie du scénario voit l'arrivée d'un seul groupe composé de quatre individus. Après deux minutes, le groupe se sépare en deux : un groupe de deux intrus et deux intrus solitaires. Plus tard, le groupe de deux se divise à nouveau. Nous avons alors un intrus pour chaque cible à protéger.

Pour ce scénario les automates, qui ont donc de très mauvaises performances (seulement 30% des intrus sont interceptés automatiquement sans opérateur), sont activés par défaut au démarrage.

4.3.4 Mesures

Les fichiers journaux sont issus de différentes parties du simulateur :
- l'environnement simulé : état des UAVs, intrus et alarmes ;
- les algorithmes : commandes générées automatiquement ;
- l'interface : l'ensemble des clics souris ainsi que les commandes résultantes.

A la fin de l'expérimentation, chaque sujet est interrogé afin d'évaluer son expérience. Les entretiens portent sur :
- les performances et prédictibilités perçues du système (pour une évaluation globale) et de chaque fonction automatique, qui sont évaluées sur une échelle de 1 à 5 ;
- la stratégie d'utilisation des sujets (leur comportement) pour chaque automate avec justifications (question ouverte).

Les entretiens nous ont permis d'estimer la confiance des sujets envers les différentes fonctions automatiques. Pour cela, nous nous sommes inspirés des travaux de Jian [54] et nous évaluons la confiance à partir du vocabulaire et des expressions françaises similaires à celui et celles, anglophones, identifiés par Jian et utilisés spontanément par les sujets durant les entretiens.

4.4 Analyse des tâches

4.4.1 Agrégation et prédiction

4.4.1.1 Opération manuelle

Pour agréger des alarmes, un opérateur utilise la souris pour sélectionner les alarmes (petit rond) et valide l'opération avec le bouton de commande correspondant. Cette tâche est plutôt difficile à réaliser : un grand nombre d'opérations manuelles sont nécessaires chaque fois qu'une alarme apparaît.

Pour prédire la trajectoire d'un intrus, un opérateur doit sélectionner l'agrégation voulue, et cliquer dans la direction désirée, puis il doit sélectionner la commande correspondante. Cette opération est facile à réaliser aussi bien cognitivement qu'ergonomiquement.

4.4.1.2 Automate

L'automate calcule s'il est vraisemblable que deux alarmes aient été déclenchées par un même intrus. Pour cela, il vérifie alors que la vitesse moyenne de l'intrus entre les deux alarmes est inférieure à la vitesse maximale d'un intrus. En effet, ces

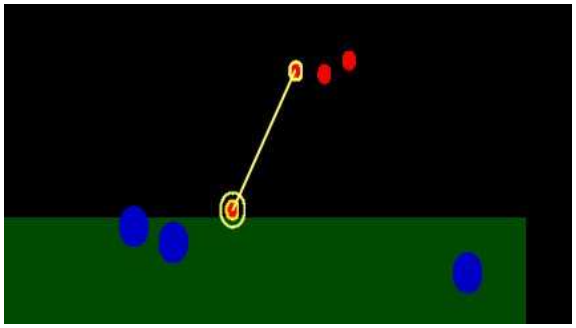

FIG. 4.4 : Pré-visualisation (en jaune) de l'agrégation de deux alarmes (rouge cerclé de jaune) avant validation de la commande.

derniers sont considérés comme des piétons qui se déplacent à une vitesse inférieurt à un dixième de la vitesse d'un UAV. Si la vitesse de l'individu est inférieurt à ce seuil, alors l'automate agrège les alarmes ensemble.

L'automate prédit la direction de la trajectoire de l'intru dans l'alignement des deux dernières alarmes d'une agrégation (de la plus ancienne vers la plus récente).

4.4.1.3 Résultats

Bien que la tâche d'agrégation ne soit pas difficile cognitivement parlant, le grand nombre d'interactions manuelles nécessaires pour la réaliser amène l'opérateur à principalement utiliser l'automate. En effet, nous observons en moyenne un taux d'utilisation des commandes manuelles (taux_{cm}) de seulement 0.48% ($\sigma = 0.01\%$).

$$\text{Taux}_{cm} = \frac{\text{Nb commande manuelle}}{\text{Nb commande manuelle} + \text{Nb commande automatique}} \tag{4.1}$$

$$\text{Taux}_{cm}\ \text{moyen} = \frac{\sum \text{Taux}_{cm_{sujet}}}{\text{Nb sujet}} \tag{4.2}$$

La prédiction manuelle est, aussi largement, ignorée par les opérateurs. Comme il est expliqué dans la section 4.2.1, il est possible de réaliser une interception manuelle sans instancier au niveau du système l'agrégation et la prédiction de trajectoire correspondante.

Dans les faits, la plupart des opérateurs ont ignoré ces deux tâches. Les entretiens nous ont permis d'obtenir quelques explications sur ce point. Ces fonctions automatiques étaient désactivées de façon permanente par certains sujets (22%), qui les considéraient comme non fiables. L'un des sujets précise même "*que [l'automate] [lui] avait relié deux alarmes qui n'avaient rien à voir*". Dans ce cas, la tâche a alors été réalisée mentalement, puisqu'il n'y a pas eu d'augmentation des commandes manuelles. D'ailleurs, un sujet précise "*les [alarmes] se suivent on voit bien*

qu'elles sont associées". Pour les deux cas concernés, tandis que l'un est ni méfiant ni confiant, l'autre est très méfiant vis-à-vis du système. Bien qu'il y ait un rejet de l'automate, la confiance des opérateurs n'est pas aussi négative.

Dans les autres cas, ces fonctions étaient activées parce qu'elles étaient nécessaire pour "*savoir dans quelle direction les ennemis allaient*". A nouveau, il n'est pas possible d'établir un lien clair avec la confiance des sujets dans le système car même si les sujets de ce groupe font majoritairement confiance, deux d'entre eux ont un avis neutre sur la question.

4.4.2 Interception

4.4.2.1 Opération manuelle

Pour effectuer les interceptions manuellement, l'opérateur a besoin de sélectionner au moins deux UAVs, et de désigner deux points entre lesquels les UAVs s'aligneront suffisamment près les uns des autres pour créer la plus longue ligne d'interception possible.

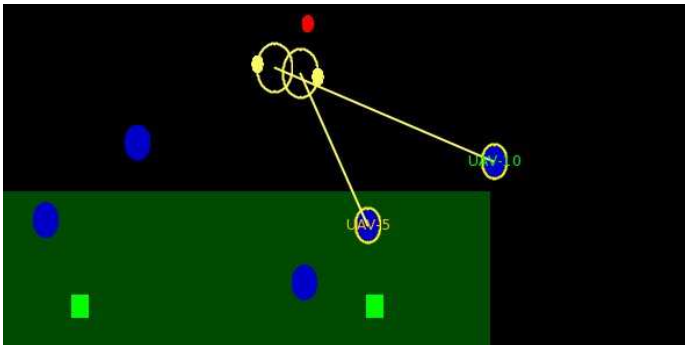

FIG. 4.5 : Pré-visualisation (en jaune) de l'interception d'un intrus à l'aide des deux UAVs sélectionnés (bleu cerclé de jaune) avant validation de la commande.

4.4.2.2 Automate

L'automate a besoin de la prédiction de trajectoire pour calculer l'emplacement de l'interception. En effet, deux UAVs seront alignés perpendiculairement à la direction suivie par l'intrus, devant de la dernière alarme, à une distance fixe.

4.4.2.3 Résultats

L'interception, contrairement aux fonctions automatiques précédentes, n'a pas seulement deux modes de fonctionnements mais en permet un troisième, intermédiaire entre le mode manuel et le mode automatique. En effet, l'opérateur a la

possibilité de corriger, ou d'améliorer les solutions générées par l'automate. L'analyse de ces commandes nous permet une interprétation en termes de confiance. Nous identifions deux catégories de commande manuelle :

- la première n'est pas corrélée aux actions de l'automate, elle n'a donc aucune influence sur l'évaluation de la confiance. Ces commandes, de par leur localisation et les UAVs sélectionnés, n'interfèrent pas avec les décisions de l'automate. L'intervention a lieu dans une zone de la carte différente de l'interception, et les UAVs opérés ne font pas partie des UAVs qui participent à l'interception ;
- la seconde catégorie inclut les actions manuelles dont la localisation est proche des actions de l'automate, ainsi que les actions qui modifient celles de l'automate (typiquement, modifier la position de l'interception). Nous interprétons cela comme un manque de confiance de l'opérateur, qui ressent le besoin de remplacer ou de compléter les interceptions réalisées par l'automate. Les commandes sont alors considérées comme correctrices.

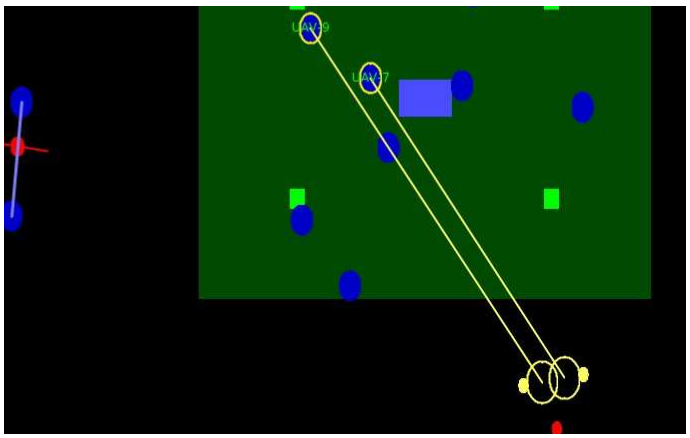

FIG. 4.6 : Exemple de commande d'interception manuelle n'interférant pas avec celle du système (tout à gauche).

Nous décidons d'étudier une possible corrélation entre le degré de confiance et le ratio du nombre de commandes de la seconde catégorie sur le nombre de commandes totales (manuelles ou automatiques) quand l'automate est actif.

$$\text{Ratio}_{correction} = \frac{\text{N command}_{catgorie2}}{\text{N command}_{catgorie1} + \text{N command}_{catgorie2} + \text{N command}_{automatique}} \tag{4.3}$$

Le résultat obtenu est illustré fig.4.8. Nous avons une corrélation de -0.63 pour le premier scénario[1] (trait plein et 'o' pour chaque échantillon[2]), et de -0.65 pour

[1] Un échantillon est considéré comme aberrant et n'a pas été pris en compte.
[2] Pourcentage de commandes correctrices en fonction du degré de confiance du sujet

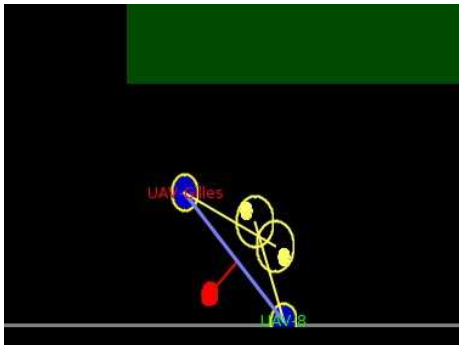

FIG. 4.7 : Exemple de commande d'interception manuelle à but correctif.

le second scénario[3] (trait pointillé et '+' pour chaque échantillon). Nous pouvons alors envisager l'existence d'une corrélation entre la confiance et le pourcentage d'actions correctives (rappelons toutefois que notre échantillon est insuffisant, le résultat n'est que qualitatif). Alors qu'il n'était pas possible d'estimer la confiance en s'appuyant sur l'interaction avec les fonctions automatiques précédentes, nous pouvons envisager, dans le contexte des interceptions, l'existence d'une corrélation entre le comportement et la confiance, due au nombre important de possibilités d'interactions.

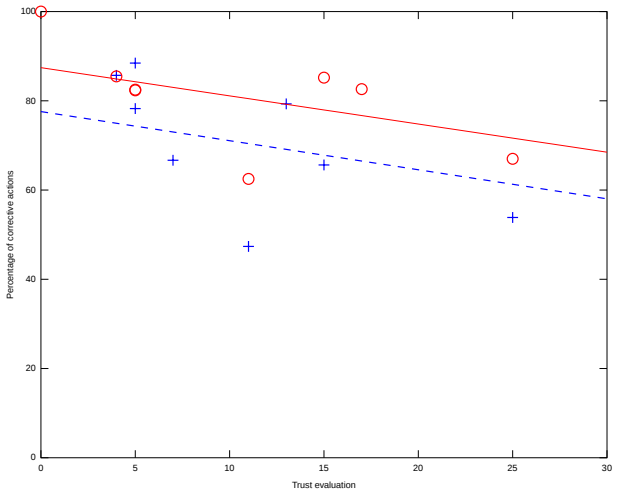

FIG. 4.8 : Corrélation entre l'évaluation de la confiance et le pourcentage de commandes manuelles qui complètent ou corrigent les décisions de l'automate. Les 'o' représentent les échantillons au cours du premier scénario, les '+' pour les échantillons au cours du second. Cœfficient de corrélation : -0.63 pour le scénario 1 (trait plein) et -0.65 pour le scénario 2 (trait pointillé).

[3]Deux échantillons sont considérés comme aberrants et n'ont pas été pris en compte.

4.4.3 Ravitaillement

Les UAVs ont une quantité limitée de carburant, et ne sont pas capables de voler sans être ravitaillés pendant le déroulement d'un scénario. Quand un UAV par se ravitailler, il se dirige vers une base de ravitaillement, où ses réservoirs sont remplis progressivement (en moins d'une minute).

4.4.3.1 Opération manuelle

L'opérateur doit sélectionner les UAVs à ravitailler, choisir une base de ravitaillement et sélectionner la commande correspondante. Les niveaux de carburant des UAVs sont apparents seulement quand ces derniers sont sélectionnés. Ainsi, pour agir manuellement, l'opérateur doit vérifier régulièrement le niveau de carburant de chaque UAV.

4.4.3.2 Automate

L'automate fait ravitailler un UAV lorsque que son niveau de carburant passe sous un seuil (10%), qui garantit un temps de vol suffisant pour rallier la base de ravitaillement quelque soit le point de départ sur la carte.

4.4.3.3 Résultats

Durant les expériences, deux schémas de comportement apparaissent :
- trois sujets n'ont réalisé aucune opération manuelle, ils ont laissé l'automate prendre en main la tâche ;
- les six autres ont laissé l'automate actif mais ont agi par anticipation en envoyant manuellement les UAVs se ravitailler avant le seuil des 10%. Pour ces sujets, les commandes manuelles de ravitaillement représentent 92% de l'ensemble des commandes de ravitaillement, soit un ou deux ravitaillements automatiques pour une moyenne de 14 au cours d'un scénario.

La figure 4.9 illustre le comportement d'un membre du second groupe de sujets. Nous observons qu'hormis une action de ravitaillement d'un drone qui a son niveau de carburant à 100%, tous les ravitaillements ont eu lieu entre 22% et 43%, ce qui est plutôt homogène. Les mesures sont en accord avec le résultat des entretiens. Ces sujets ont déclaré qu'ils estimaient le seuil de déclenchement du ravitaillement trop bas et qu'ils se seraient sentis en sécurité avec un seuil plus élevé.

Nous pouvons conclure que le premier groupe de sujets se fiait complètement (faisait confiance) à la fonction de ravitaillement automatique proposée, et que le second groupe, plutôt qu'être considéré comme méfiant, pensait simplement que le système n'était pas configuré avec un seuil suffisant. La fonction de ravitaillement automatique avait un comportement trop simple pour que la notion de confiance puisse être pertinente afin de déterminer l'usage ou le non-usage de l'automate.

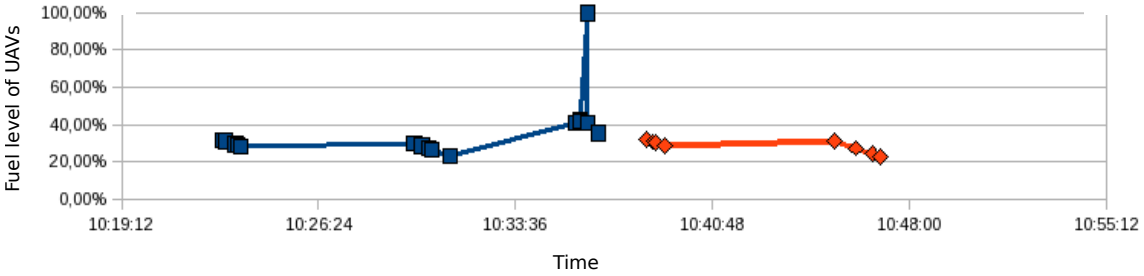

FIG. 4.9 : Niveau de fuel des UAVs lorsqu'ils reçoivent une commande de ravitaillement.

4.5 CONCLUSIONS

Les travaux précédents ont permis une première analyse du comportement d'un opérateur. En cela, ils valident notre approche de l'évaluation de la confiance par l'observation des interactions. En effet, pour des fonctions, telles que l'interception, qui ont besoin à la fois d'une interaction complexe et d'un traitement mental complexe, nous avons identifié une corrélation claire entre la confiance et le taux d'interactions.

De plus, un second point est soutenu par cette expérimentation. Celle-ci met en évidence qu'un système n'est pas perçu dans sa globalité mais à un niveau de granularité plus fin, ce qui est en accord avec la définition de l'objet de confiance que nous donnons en introduction du chapitre 2. En effet, les opérateurs distinguent les différentes fonctionnalités d'un système et interagissent différemment avec ces dernières.

Cette première étude n'est qu'exploratoire et n'a pour but que de valider les principes de notre approche. La confiance n'est pas systématiquement impliquée dans toutes les fonctions automatiques que nous avons étudiées. Ce point souligne la difficulté de relier les observables et les comportements de confiance, d'autant plus que nous nous attendions à obtenir des informations à partir des fonctions qui supposent une interaction complexe (prévision de trajectoire, agrégation géographique ou topologique de cibles élémentaires), alors que nous avons manqué de signes explicites de confiance.

Nous avons abordé dans l'état de l'art toute une partie sur la modélisation du dialogue qui n'est pas du tout exploitée ici. Ainsi, il est nécessaire de renouveler l'expérience en la basant, cette fois, sur un modèle de dialogue que nous aurons défini préalablement. Celui-ci doit permettre de cibler précisément nos observables en mettant en avant la notion de contrôle. Ainsi des signes explicites de confiance seront rendus observables.

5 La confiance au sein du dialogue

Dans notre état de l'art sur la confiance et le contrôle supervisé, nous avons fortement insisté sur la nécessité d'évaluer la confiance d'un opérateur envers son système. Nous avons ensuite envisagé de pouvoir faire reposer cette évaluation sur le dialogue qui s'établit entre l'homme et la machine. Mais pour cela, il nous faut d'abord parler du lien entre confiance et dialogue. Nous proposerons ensuite quelques modifications représentatives du modèle de contrôle supervisé avant d'y intégrer notre modèle de dialogue.

5.0.1 Confiance et contrôle

Nous avons introduit la notion de confiance au chapitre 2 de ce manuscrit. Voyons maintenant comment cette notion est mise en relation avec la notion de contrôle. Nous partons de l'hypothèse que, de la même manière que la notion de confiance dans les relations interpersonnelles peut être appliquée aux relations Hommes-machines [54], la relation entre confiance et contrôle décrite dans le cadre des relations interpersonnelles est applicable aux relations Hommes-machines.

La notion de contrôle est définie par Green et Welsh comme "un processus cybernétique régulatoire qui dirige ou contraint une activité interactive à une norme ou un but"[1]. A partir de cette définition, Das et Teng [36] considèrent que le contrôle a pour rôle la surveillance d'une tâche pour sa bonne réalisation. Il est à noter que Castelfranchi considère le contrôle non seulement comme un processus de surveillance mais aussi comme un processus d'intervention afin que tout écart de la tâche vis-à-vis de l'objectif soit corrigé. Pour Castelfranchi l'action de contrôle [16] :

- "vise à déterminer si une autre action a été exécutée avec succès ou si un état donné du monde a été atteint ou maintenu (vérification, information de retour)[2]". Autrement dit, le contrôle est un processus de consultation de l'information. Un opérateur lit un nombre plus ou moins important d'informations afin de contrôler l'activité du système ;

[1] a cybernetic, regulatory process that directs or constrains an interactive activity to some standard or purpose

[2] *aimed at ascertaining whether another action has been successfully executed or if a given state of the world has been realized or maintained (feedback or checking)*

– "vise à traiter les écarts éventuels et les événements imprévus afin d'y faire face correctement (intervention)[3]". Pour cela l'opérateur traite et manipule les informations qu'il a acquises du système afin d'élaborer les commandes qu'il communiquera au système.

La relation entre confiance et contrôle est complexe, et les recherches sur ce thème proposent des interprétations contradictoires [37]. Les deux points de vue principaux sont :

– la substitution : la confiance se substitue au contrôle. Nous avons donc une relation d'opposition entre les deux [39]. Lorsque le degré de confiance est faible, celui du contrôle est élevé. Et inversement. Cette approche se base sur une analyse économique de la confiance. La confiance incite à la coopération, réduit l'incertitude et augmente les échanges d'informations [48]. Ainsi, plus la confiance est élevée plus le coût de la surveillance et d'autres mécanismes de contrôle est faible.
– la complémentarité : la confiance et le contrôle se renforcent mutuellement [37]. Les mécanismes de contrôle fournissent des règles et des mesures objectives sur lesquelles baser une décision de confiance.

Ces deux points de vue n'ont pas été départagés. Certains travaux montrent que le premier point de vue s'applique très bien à un contexte coopératif [35] tandis que le second s'applique plutôt à un contexte hiérarchique [8].

En ce qui nous concerne, nous sommes dans une relation de coopération. Nous nous plaçons donc, a priori, dans le premier groupe où confiance et contrôle s'opposent. Cela signifie que le processus de consultation de l'information prend une place de plus en plus importante dans l'interaction homme-machine au fur et à mesure que la méfiance de l'opérateur se développe. Or nous avons vu, dans la littérature, une approche du dialogue centrée sur l'information et fondée sur la théorie du grounding. Cette théorie peut nous permettre de définir un modèle de dialogue qui modélise ce processus de consultation de l'information, et ainsi nous pouvons analyser la confiance d'un opérateur par l'observation des contrôles exercés par celui-ci sur le système. L'évaluation de la confiance est alors fondée sur une analyse des mécanismes de contrôle mis en œuvre par un opérateur au cours du dialogue avec un système.

5.1 Contrôle supervisé

Rappelons que Sheridan [102] considère le contrôle supervisé comme un ensemble de cinq fonctions reliées les unes aux autres (fig.1.4). Dans ce modèle, l'intervention de l'opérateur peut correspondre à :

1. l'ajustement d'une consigne (ou commande) existante : fonction d'intervention du contrôle supervisé,

[3] *aimed at dealing with the possible deviations and unforeseen events in order to positively cope with them (intervention)*

2. à la génération de nouvelles consignes (ou commandes) : fonction de programmation du contrôle supervisé,
3. à une intervention humaine directement sur la tâche (par exemple, enlever une bouteille renversée sur une chaîne d'embouteillage) : fonction d'intervention du contrôle supervisé.

On peut dans un premier temps considérer que la fonction intervention est fortement assimilable à la fonction programmation. En effet, les deux ont pour objectif la saisie d'instructions pour la réalisation d'une tâche. Bien sûr les motivations divergent selon les deux cas. En effet, la fonction de programmation n'est motivée que par la réalisation de nouvelles tâches tandis que la fonction d'intervention est motivé par le besoin de modifier des instructions en cours suite à un incident ou pour améliorer la réalisation de la tâche. C'est pourquoi Sheridan définit un rebouclage entre la sortie de la fonction intervention et l'entrée de la fonction programmation. L'intervention humaine (autre que par commande(s) et/ou consigne(s)) est considérée comme une sortie de script, une exception. **Nous considérons que les interventions à but correctif contiennent une part d'information concernant la confiance d'un opérateur**. En effet, dans la section 5.0.1, nous avons vu que la notion de contrôle possède deux aspects dont l'intervention. C'est pourquoi nous devons définir un modèle de dialogue qui nous permette d'analyser ces interventions. Bien entendu il faudra différencier les actions de l'opérateur liées à la fonction de programmation et celles liées à la fonction d'intervention. Nous pensons que cette différence s'établira au travers de la façon d'agir de l'opérateur.

La fonction d'apprentissage amène l'opérateur à redéfinir, à étendre ou à synthétiser ses objectifs et les solutions allant avec. D'où le rebouclage de la fonction Learn vers la fonction Plan.

Nous pouvons donc considérer le contrôle supervisé sur trois niveaux (voir fig.5.1) qui correspondent aux trois rebouclages du modèle :

Planification/Préférence : apprentissage d'un modèle mental du système par l'opérateur et application de celui-ci à la définition des objectifs et des solutions ;

Commande/Consigne : génération des commandes et consignes, non seulement initiales mais aussi lors des interventions de l'opérateur ;

Suivi : surveillance de l'état du système et de l'environnement afin de s'assurer que les objectifs sont atteints.

Cette vision du contrôle supervisé met en évidence deux niveaux de dialogue entre un opérateur et le système.

Le premier est le suivi. En effet, le système informe en continu l'opérateur de son état, de celui de la tâche et de l'environnement. Or le partage d'informations par le système engendre un dialogue qui permet un contrôle de l'information par l'opérateur. En accord avec sa définition donnée à la section 5.0.1, **le contrôle peut donc être ramené à un processus de consultation de l'information**.

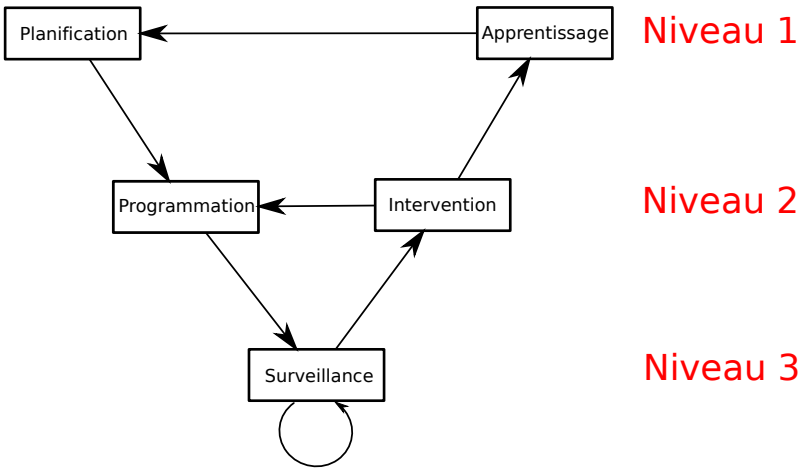

FIG. 5.1 : Découpe du contrôle supervisé en trois couches

En effet, de façon très basique le simple fait d'acquitter l'information (valider un message ponctuel, focaliser son attention visuelle sur un objet, etc.) exprime un contrôle de l'opérateur sur le système. Ainsi l'observation de ce contrôle peut nous renseigner sur la confiance de l'opérateur. Il nous faut donc aborder la modélisation du dialogue au niveau de la surveillance.

Le second niveau de dialogue concerne l'intervention d'un opérateur sur le système. Il s'intéresse à la manière dont l'opérateur instruit le système. Il est intéressant d'observer comment la contribution d'un système à l'élaboration des commandes est acceptée ou rejetée au sein du dialogue. En effet, l'aide du système peut-être à ce moment plus ou moins contrôlé par l'opérateur. De plus comme nous l'avons dit plus haut, l'intervention de l'opérateur est représentative de la notion de contrôle définie à la section 5.0.1. **Le contrôle n'est plus seulement un processus de consultation de l'information mais aussi de manipulation**. En effet, l'opérateur agit indirectement sur les informations retournées par le système en modifiant son activité.

Ainsi, afin d'anticiper sur des situations où un opérateur utilise mal le système par manque de confiance, voire le désactive avec toutes les conséquences que cela peut avoir (section 1.5.3), il est nécessaire d'évaluer la confiance. Pour cela nous avons proposé à la section 2.5.3 une approche basée sur l'analyse du dialogue. Nous avons par ailleurs établi un lien entre champ commun et confiance au travers de la notion de contrôle. C'est pourquoi nous proposons d'implémenter un gestionnaire de dialogue basé sur la théorie du grounding qui puisse analyser ces deux niveaux du contrôle supervisé.

5.2 Système de dialogue

5.2.1 Adaptation du modèle de Traum au contrôle supervisé : Monitoring

Nous allons nous focaliser sur l'une des cinq fonctions du contrôle supervisé : la surveillance, à savoir la troisième couche du contrôle supervisé qui est définie comme le suivi de l'état du système et de l'environnement afin de s'assurer que les objectifs sont atteints. Nous proposons ici d'adapter le modèle de Traum présenté au troisième chapitre d'une part, au regard des besoins de la fonction de surveillance, et d'autre part, par la mise en avant des mécanismes de contrôle. En effet, notre évaluation de la confiance est basée sur le lien confiance-contrôle.

5.2.1.1 Information continue ou événementielle

La première question qui se pose est "comment, lors du monitoring, les informations sont elles présentées à l'opérateur au cours du temps ?". En effet, nous observerons que, selon l'accessibilité de l'information, le dialogue ne se construit pas de la même manière.

On distingue pour cela trois modes d'accessibilité :

Accessibilité continue : l'information est présentée en permanence par le système. Ce dernier se contente de l'actualiser régulièrement. En terme de grounding cela revient à la dynamique suivante :

1. I vers 1 : initialisation de l'unité de dialogue par le système,
2. 1 vers 1 : correction de l'information par le système,
3. 1 vers F : acquittement de l'opérateur,
4. F vers 1 : réouverture de l'unité de dialogue par le système en raison d'une correction de l'information.

Dans ce cas, le modèle de grounding de Traum se simplifie grâce à la suppression d'un grand nombre de transitions (fig.5.2) :

- l'opérateur ne corrige pas les informations du système. En effet, du point de vue du contrôle supervisé, nous nous situons dans la couche de suivi ce qui signifie que l'opérateur surveille l'état du système et de l'environnement : il ne fait que s'informer. Les transitions "Repair(R)" et "ReqRepair(I)" ne sont plus nécessaires ;
- une unité de dialogue correspond à une information précise. Bien que le système puisse maintenir en parallèle plusieurs unités de dialogue, il ne transmet pas plusieurs informations différentes au sein d'une même unité. Ainsi, les transitions "continue(I)" du modèle de Traum ne sont pas utiles ;
- enfin, l'unité de dialogue étant continuellement actualisée, elle ne peut pas devenir inactive : l'état D est alors inutile ainsi que les transitions qui y mènent.

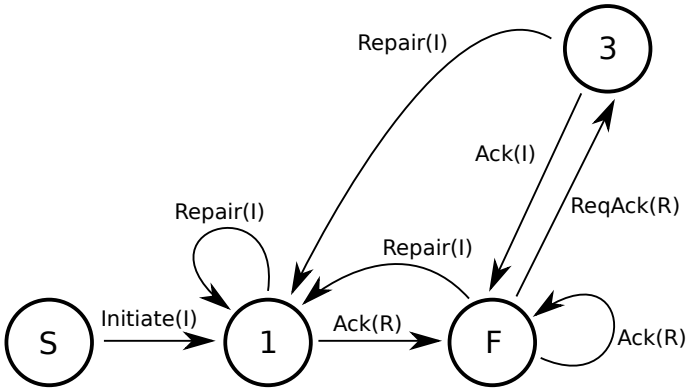

FIG. 5.2 : Réseau de transitions pour le monitoring d'informations continues

Accessibilité discrète : une information ponctuelle, contrairement à l'information continue, se clôt définitivement. De plus, l'état D reste nécessaire, car une information peut devenir, avec le temps, inutile même si le répondant ne l'acquitte pas. Dans ce cas, l'initiateur peut mettre fin à l'unité de dialogue en cours (fig.5.3).

Accessibilité discrète sur requête : l'élément supplémentaire est l'inversion des rôles. En effet, l'opérateur devient l'initiateur et le système le répondant. Afin d'avoir une modélisation la plus simple possible, nous considérerons que la requête de l'opérateur forme à elle seule une unité de dialogue, dont l'acquittement par le système correspond à l'initialisation d'une nouvelle unité de dialogue. Cette dernière a, pour initiateur, le système qui présente l'information demandée et a, pour répondant, l'opérateur. Nous nous retrouvons alors exactement dans la même situation que pour une accessibilité discrète (sans requête) et le même modèle de dialogue peut donc être utilisé.

Par exemple, un opérateur de drone souhaite accéder à des informations météorologiques. Pour cela, lorsqu'il formule sa demande, une unité de dialogue est initiée. L'information "demande d'information météo" est acquittée par le système avec l'affichage de la météo. Ceci clôt l'unité de dialogue initié par l'opérateur mais en ouvre une seconde. Cette seconde unité de dialogue porte sur le partage de l'information "météo" par le système. Elle suivra alors le même modèle que pour une information à accessibilité discrète. Elle prendra alors fin :

- soit avec l'acquittement par l'opérateur de l'information,
- soit lorsque le système abandonnera le dialogue (si l'acquittement n'est toujours pas formulé après un certain laps de temps).

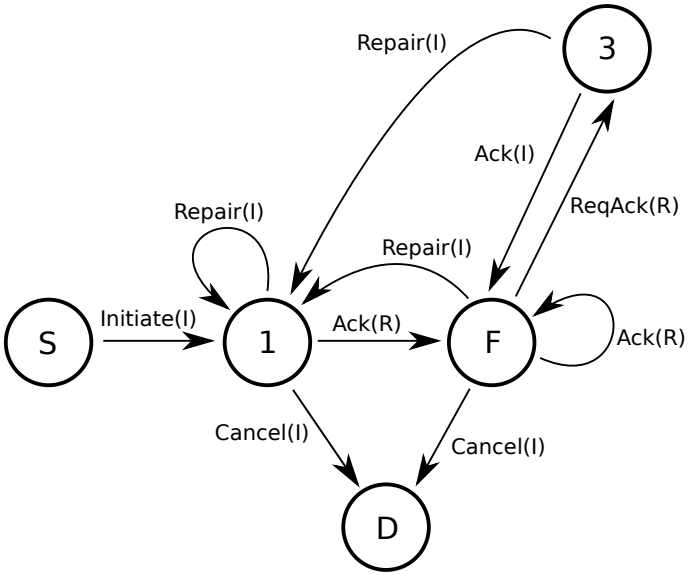

FIG. 5.3 : Réseau de transitions pour le monitoring d'informations discrètes

5.2.1.2 Contrôle et confiance

Dans ce modèle de dialogue, il n'existe qu'un seul mécanisme de contrôle : l'acquittement. En effet, l'opérateur surveille l'activité du système, et le fait de focaliser son attention sur une information pour la lire est perçu comme un acquittement à l'introduction de l'information. L'observation du focus attentionnel montre que l'opérateur a pris connaissance de l'information. Ce n'est pas la seule forme d'acquittement mais plutôt la moins évidente en comparaison, par exemple, à une fonction de validation attachée à l'information.

La validation passive de l'opérateur, liée à la fonction de surveillance du contrôle supervisé, ne permet pas une analyse de l'acquittement au regard des travaux de Roque [96]. En effet, dans l'état actuel une information introduite ne peut avoir en terme de grounding que deux états : acquitté et non acquitté. Il n'y a pas de possibilité pour un acquittement avec renforcement, et encore moins d'acquittement du contenu.

En gardant en mémoire notre hypothèse que confiance et contrôle sont liées par une relation d'opposition, on propose alors de distinguer deux types de comportements liés à la confiance : le premier, un comportement de confiance, où l'opérateur n'effectue pas de contrôle, c'est-à-dire qu'il n'acquitte pas les informations présentées par le système, et le second, un comportement de méfiance, où il effectue un contrôle de l'information en portant son attention sur celle-ci. Dans ce dernier cas, il est envisageable de relativiser le degré de confiance d'un opérateur à une fréquence de contrôle de l'information comparable à la mesure de confiance objective de Freedy [47].

5.2.1.3 Au-delà du grounding : extension, explication

Une forme de contrôle qui n'est pas explicite dans le modèle de grounding de Traum, et donc dans notre version adaptée à la fonction de surveillance, est la comparaison des informations et la recherche d'explication par l'opérateur. Nous considérons ces deux mécanismes comme des mécanismes de contrôle très importants et donc à mettre en exergue au sein du modèle. C'est pourquoi nous avons décidé d'ajouter un certain nombre de transitions au modèle de Traum. Cela implique l'ajout de deux mécanismes :

- l'extension de l'information : le répondant demande des informations complémentaires (par exemple des précisions, des compléments) lors de la lecture d'une information. Ces éléments renforcent le degré de grounding de l'information initiale, mais surtout – et c'est ce qui nous intéresse – introduisent un mécanisme complémentaire de contrôle de l'information. Ainsi, des transitions ReqExt(R) – pour la demande d'informations complémentaires – et Ext(I) – pour la complétion des informations – sont introduites au sein du réseau de transitions ;
- l'explication : le répondant demande une explication, une raison à l'information consultée. Il demande en quelque sorte "pourquoi ?, comment ?". Il est alors nécessaire d'adjoindre des transitions ReqExpl(I) – pour les demandes d'explications – et Expl(I) – pour les explications – au modèle de Traum.

A priori, ces mécanismes sont mis en jeu par l'opérateur. Par conséquent, dans le modèle de Traum adapté au monitoring, seul le répondant les déclenche. Cela signifie que les transitions ReqExt(R) et ReqExpl(R) partent de l'état 1 pour rejoindre l'état 2 (voir fig.5.5). A la réponse du système correspondent les transitions Ext(I) et Expl(I), qui font passer le réseau de transitions de l'état 2 à l'état 1.

Ces mécanismes peuvent aussi être initiés à partir de l'état F. L'unité de dialogue a déjà été close par un acquittement de l'opérateur, mais ce dernier exprime le besoin d'obtenir ces informations malgré tout. Ainsi, des transitions ReqExt(R) et ReqExpl(R) relient l'état F à l'état 2.

Les actions décrites précédemment supposent, *a priori*, que seul l'opérateur déclenche les mécanismes d'explication et de complément informatif par des requêtes. Mais le système peut, de lui-même, donner ces explications et/ou compléments d'informations. Ainsi, des transitions Ext(I) et Expl(I) relient les états 1 et F à l'état 1.

Enfin la mise en œuvre de ces mécanismes permet cette fois l'observation d'acquittement renforcé. En effet, lors de l'acquittement, il est possible, pour l'opérateur, de poursuivre par un acte de langage qui se traduira modèle par une transition de demande d'explications ou d'informations étendues dans notre. On acquiert ainsi un nouveau degré de grounding au sein de notre modèle. Cela se traduit donc par un troisième comportement qui, cette fois, peut traduire la méfiance de l'opérateur face à l'information.

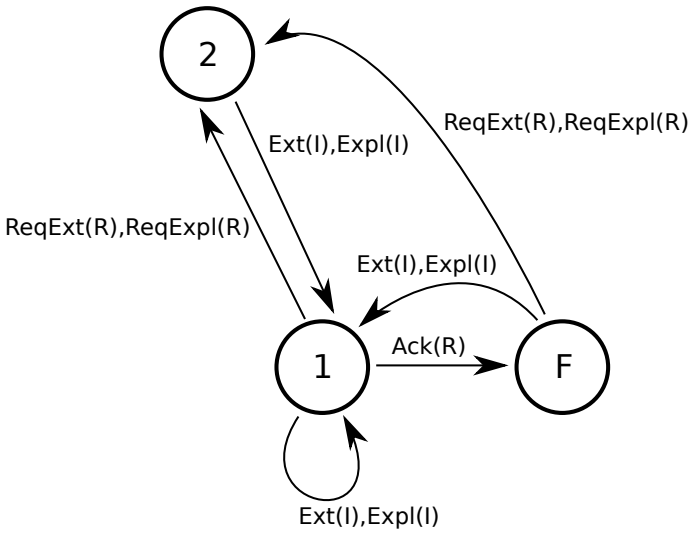

FIG. 5.4 : Complément du réseau de transitions pour les mécanismes d'explication et de complément informatif

5.2.1.4 Exemple : régulateur de vitesse

Un régulateur de vitesse a pour objectif de maintenir la vitesse d'un véhicule au niveau de la consigne donnée par le conducteur. Généralement, seule la vitesse instantanée est affichée et consultable en continu. Regardons comment en modifiant l'interaction avec ce régulateur nous pouvons appliquer notre modèle et ainsi évaluer la confiance du conducteur.

D'abord nous ajoutons un outil de suivi du regard. Ainsi le régulateur (illustré fig.6.8.3) saurait quand le conducteur consulte l'information de vitesse.

Ensuite notre régulateur de vitesse, en plus d'afficher la vitesse instantanée, se verrait doté de deux fonctions supplémentaires. Ainsi, le conducteur pourrait demander, lorsqu'il le souhaite, des informations complémentaires à propos de la vitesse et ainsi contrôler la cohérence de celle-ci avec d'autres informations (par exemple régime moteur et déclivité de la route) de façon temporaire. Il pourrait également demander des explications sur la vitesse qui est affichée. A ce moment, le régulateur de vitesse pourrait formuler deux réponses :

- la vitesse est obtenue par tachymétrie
- le conducteur est en train d'agir sur l'accélérateur.

Le conducteur pourrait donc interagir avec le régulateur en mettant en œuvre des mécanismes de contrôle qui seraient, selon nous, révélateur de la confiance qu'il porterait à son régulateur.

Si nous nous reportons à notre modèle de dialogue, nous avons :

- une information présentée en continu : à chaque acquittement par focus attentionel, le réseau de transition passe de l'état 1 à l'état F.

FIG. 5.5 : Exemple d'interface pour un régulateur de vitesse. La vitesse instantanée est affichée en vert. On retrouve, sur le côté droit, trois boutons qui permettent (de haut en bas) d'accéder à des informations complémentaires sur la vitesse, de demander une explication et de saisir une consigne. Les informations complémentaires et les explications viendraient s'afficher en lieu et place de la vitesse de façon temporaire.

Lors du rafraîchissement de l'information par le régulateur de vitesse, le réseau de transition revient à l'état 1 (transition Repair(I) à partir de l'état F si il y a eu acquittement, ou de l'état 1 autrement).

– une extension de l'information : sur demande du conducteur qui déclenche la transition ReqExt(R) de l'état F à l'état 2. L'affichage des informations complémentaires déclenche la transition Ext(I) et ramène le réseau à l'état 1. Le timeout se résout simplement par une transition repair(I).
– une explication : sur demande du conducteur qui déclenche la transition ReqExpl(R) de l'état F à l'état 2. La réponse est sélectionnée selon le contexte, et son affichage déclenche la transition Expl(I) de l'état 2 vers 1.

Ainsi selon les interactions mises en œuvre par le conducteur, nous pouvons évaluer son degré de confiance. En effet, selon nos hypothèses, l'acquittement par focus attentionnel, selon la fréquence à laquelle il a lieu, sera indicatif d'une faible remise en cause de la confiance du conducteur. Par contre les demandes d'explication ou d'extension de l'information seront l'expression d'un doute vis-à-vis du régulateur, et représenterons donc une certaine méfiance du conducteur.

On voit donc, à travers cet exemple, qu'il est possible, en adaptant les interactions d'un système de façon à permettre une mise en œuvre de notre modèle de dialogue, d'évaluer la confiance d'un utilisateur en fonction de ses interactions avec le système au niveau de la couche de suivi du contrôle supervisé.

5.2.2 Adaptation du modèle de Traum au contrôle supervisé : Teach/Intervene

De même que pour la couche de suivi, il est possible d'adapter le modèle de Traum à la couche de commandes/consignes. Cette couche regroupe deux fonctions du contrôle supervisé : programmation et intervention qui se définissent comme l'élaboration initiale d'instructions ou leur modification afin d'améliorer le processus de réalisation ou de corriger des erreurs. Pour ce faire nous distinguons deux cas différents :

– les actions initiées par le système que l'opérateur doit corriger ou non avant de valider ;
– les actions initiées par l'opérateur auxquelles le système peut apporter une contribution.

De même que pour la couche de suivi, nous allons voir aussi quels sont les mécanismes de contrôle qui existent au sein du modèle et comment nous pouvons les enrichir. En effet, rappelons que notre but est d'évaluer la confiance grâce à l'observation des mécanismes de contrôle mis en œuvre au cours du dialogue.

5.2.2.1 Commande initiée par le système

Si l'on se réfère aux degrés d'automatisation définis par Sheridan [101], des niveaux 6 à 9, le système entreprend de manière autonome un certain nombre d'actions en informant plus ou moins l'opérateur qui ne pourra intervenir qu'en initiant ses propres commandes. Des niveaux 2 à 5, l'automate sollicite l'intervention de l'opérateur avant d'agir. Dans tout ces cas, le dialogue résultant en terme de grounding, revient à :

1. I vers 1 : le système initialise l'unité de dialogue, la commande est présentée à l'opérateur ;
2. 1 vers 3 : l'opérateur apporte une modification à la commande ;
3. 3 vers 3 : pour chaque modification supplémentaire ;
4. 3 vers F : acquittement des modifications par le système ;
5. F vers F : acquittement de l'opérateur (optionnel).

Par ailleurs, il faut aussi considérer la variante suivante :

1. 1 vers 3 : l'opérateur apporte une modification à la commande ;
2. 3 vers F : acquittement des modifications par le système ;
3. F vers 2 : l'opérateur demande de réévaluer la commande à partir des modifications apportées ;
4. 2 vers 1 : réévaluation de la commande par le système ;
5. 1 vers F : acquittement de l'opérateur.

Ainsi ne sont plus nécessaires les transitions :

– ReqRepair(R) à partir de l'état 1, qui permet, au sein du modèle de Traum, au répondant de demander à l'initiateur de répéter et/ou de corriger l'information présentée. Dans notre cas, cette possibilité devient inutile. En effet, si aucune modification n'a lieu, une demande de réévaluation de la commande n'apporte rien. Elle est nécessaire uniquement si l'opérateur modifie les paramètres de la commande. Or, dans ce cas, la transition ReqRepair(R) a lieu à partir de l'état 2. Nous pourrions supposer que le délai avant l'activation de la transition ReqRepair(R) est suffisamment long pour que l'évolution de l'environnement

soit significative. Ainsi, la demande de réévaluation pourrait aboutir à une commande différente. Mais nous considérons dans ce cas que l'unité de dialogue est déjà passée à l'état D, du fait du temps écoulé sans réponse du répondant.

– ReqRepair(I) qui permet à l'initiateur de demander au répondant de corriger sa proposition. Elle devient vaine. En effet, le système peut demander une validation auprès de l'opérateur (un acquittement) avant d'exécuter une commande, ou quand bien même l'opérateur souhaiterait appliquer des modifications, le système ne requérira pas une modification de la commande qu'il vient de calculer.

Enfin, de la même manière que pour le monitoring, des transitions pour les demandes d'explications ou de complément informatif sont à ajouter dans le modèle.

Nous obtenons alors le tableau de transitions de tab.5.1.

Acte suivant	Etat courant						
	S	1	2	3	*4*	F	D
Initiate	1						
Continue(I)		1			*4*		
Continue(R)			2	3			
Repair(I)		1	1	1	*4*	1	
Repair(R)		3	2	3	*3*	3	
ReqRepair(I)			*4*	*4*	*4*	*4*	
ReqRepair(R)		*2*	2	2	*2*	2	
Ack(I)				F	*1**	F	
Ack(R)		F	F*			F	
ReqAck(I)		1				1	
ReqAck(R)				3		3	
Cancel(I)		D	D	D	*D*	D	
Cancel(R)			1	1		D	
Explain(I)		**1**	**1**			**1**	
Extend(I)		**1**	**1**			**1**	
ReqExplain(R)		**2**				**2**	
ReqExtend(R)		**2**				**2**	

TAB. 5.1 : Réseau de transitions spécifié pour les commandes du systèmes avec intervention humaine. En *italique* les transitions supprimées du modèle de Traum, ainsi que l'état 4 n'ayant plus aucune transitions y conduisant. Et en **gras** les transitions ajoutées au modèle de Traum.

A nouveau nous avons introduit au sein du modèle les mécanismes d'extension de l'information et d'explication. Ces mécanismes portent en terme de contrôle la même significativité que lors du suivi. Ils sont donc des indicateurs d'une remise en cause des propositions ou actions de l'automate, et peuvent donc être interprétés comme une méfiance de l'opérateur vis-à-vis de l'activité du système.

5.2.2.2 Commande initiée par l'opérateur

Lors du monitoring, certaines conditions peuvent amener une intervention de l'opérateur. En effet, ce dernier peut considérer (à tort ou à raison) que l'automate est inapte face aux conditions environnementales, ou bien peut vouloir compléter les actions du système. Dans tous les cas, l'intervention ($2^{ème}$ couche du contrôle supervisé) se traduit par une unité de dialogue à l'initiative de l'opérateur, afin d'établir une commande/une consigne. Suivant le modèle de Traum, nous aboutissons typiquement à la séquence suivante :

1. I vers 1 : initialisation du dialogue par l'opérateur ;
2. 1 vers 1 : pour chaque élément de la commande, une transition "continue(I)" ;
3. 1 vers F : acquittement du système.

De plus, un certain nombre de cas où le système émet des propositions au fur et à mesure de la saisie de l'opérateur peuvent être envisagé. L'intervention du système, quelle qu'elle soit, donne la variante suivante :

1. I vers 1 : initialisation du dialogue par l'opérateur ;
2. 1 vers 3 : proposition du système pour un paramètre de la commande encore non saisi, ou une contre-proposition au paramètre saisi par l'opérateur ;
3. 3 vers F : validation de la proposition par l'opérateur ;
4. F vers F : acquittement du système.

Enfin, l'opérateur peut demander au système de compléter la commande. Il est alors nécessaire d'ajouter une transition ReqRepair(I)de l'état 1 vers l'état 4.

A nouveau, nous superposons au modèle de dialogue spécifié les mécanismes d'explication. Nous obtenons alors le tableau de transitions de la table tab.5.2.

Dans ce dernier cas, les mécanismes sont à nouveau liés à l'introduction des mécanismes d'explication et d'extension de l'information. Nous supposons que la confiance sera donc inversement proportionnelle aux quantités d'informations complémentaires et au nombre d'explications qui seront demandées par l'opérateur.

De plus, au vu de l'expérience présentée au chapitre 4, il est à noter que l'initialisation d'un tel dialogue revient à effectuer un contrôle, non pas au sens de vérification, mais au sens d'une action corrective comme le définit Castelfranchi [16]. Bien entendu, comme pour l'expérience du chapitre 4, il est nécessaire de distinguer les interventions qui initient des tâches, de celles qui corrigent un processus de réalisation.

Acte suivant	Etat courant						
	S	1	*2*	3	4	F	D
Initiate	1						
Continue(I)		1			4		
Continue(R)			*2*	3			
Repair(I)		1	*1*	1	4	1	
Repair(R)		3	*2*	3	3	3	
ReqRepair(I)		**4**	*4*	*4*	*4*	*4*	
ReqRepair(R)		*2*	*2*	*2*	*2*	*2*	
Ack(I)				F	1*	F	
Ack(R)		F	*F**			F	
ReqAck(I)		1				1	
ReqAck(R)				3		3	
Cancel(I)		D	D	D	D	D	
Cancel(R)			*1*	1		D	
Explain(R)				**3**	**3**	**3**	
Extend(R)				**3**	**3**	**3**	
ReqExplain(I)				**4**		**4**	
ReqExtend(I)				**4**		**4**	

TAB. 5.2 : Réseau de transitions spécifié pour les commandes de l'opérateur. En *italique* les transitions supprimées du modèle de Traum, ainsi que l'état 2 n'ayant plus aucune transition y conduisant ; en **gras** les transitions ajoutées au modèle de Traum.

5.3 REPRÉSENTATION DE L'INFORMATION

Précédemment, nous avons vu comment spécifier et étendre le modèle de grounding de Traum pour analyser les mécanismes du contrôle de l'information au cours du dialogue et ainsi avoir des indicateurs de confiance au sein du modèle de dialogue. Au-delà du grounding usuel, il a été nécessaire de mettre en œuvre des mécanismes pour l'obtention d'explications (justifications) et pour la demande de complétion (informations complémentaires). Ces mécanismes présupposent que le système soit capable, au cours du dialogue avec l'opérateur, d'établir des liens entre les différentes informations qu'il manipule pour étendre une information avec des informations annexes, ou pour expliquer une information. Il faut alors définir, en appoint du modèle de dialogue, un modèle de représentation de l'information qui établisse ces liens et permette ainsi de générer les réponses adéquates.

5.3.1 Types d'informations

Nous proposons plusieurs types d'informations :

alarme : information événementielle concernant l'état du système ou de l'environnement. Dans notre modèle de dialogue, elle est gérée par la couche suivi et elle a une accessibilité discrète ;

statut : information présentée en continu ou sur demande concernant l'état du système ou de l'environnement. Gérée également par la couche suivi de notre modèle, elle a une accessibilité continue ou discrète sur demande ;

retour d'activité : information concernant l'activité du système, les consignes actives ainsi que les commandes en cours. Elle est gérée par la couche suivi de notre modèle et a une accessibilité discrète ;

règle : information concernant le fonctionnement du système, c'est-à-dire les procédures que suit le système. Elle a une accessibilité discrète sur demande et ne peut s'obtenir qu'avec des demandes d'explication.

Il est nécessaire pour l'application de notre modèle de dialogue d'identifier toutes les informations manipulées par le système afin d'établir les liens nécessaires pour les mécanismes d'extension de l'information et d'explication.

5.3.2 Relation entre informations

Deux relations sont donc à définir pour établir une représentation de l'information : l'explication et l'extension d'information.

5.3.2.1 Graphe d'explication

Prenons l'exemple d'une information de cap (direction) d'un UAV, ayant un faible niveau de carburant, qui doit rejoindre un waypoint. Nous proposons de représenter les informations sous forme d'un graphe orienté. A chaque nœud du graphe correspondra une information. Les arêtes orientées signifient "explique" (fig.5.6). Ainsi nous pouvons lire : la commande "aller à" "explique" le "cap".

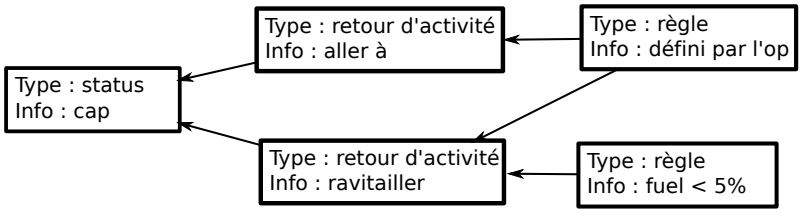

FIG. 5.6 : Exemple d'un graphe d'explication

Cette approche de la représentation de l'information au sein du dialogue a déjà été mise en œuvre dans le cadre des grammaires avec la notion de frame définie par

Minsky [75] (section 3.1.1). En effet, une frame est assimilable à un formulaire vide dont les emplacements permettent de décrire un objet en rattachant d'autres types d'informations. Ainsi dans notre cas, la frame qui définit une information possède deux emplacements terminaux pour décrire l'information : le type d'informations (présenté ci-dessus), et "l'info" à savoir le contenu de l'information en lui-même (voir fig.5.6). Enfin un troisième emplacement est nécessaire pour établir le lien explicatif. On constate que dans notre exemple, ce dernier terminal nécessite d'être défini judicieusement. Par exemple, la commande "ravitailler" peut s'expliquer de différentes manières : avec une règle de l'automate, ou par l'opérateur.

5.3.2.2 Graphe d'extension de l'information

Avec le même exemple, voyons à quoi ressemble le graphe d'extension de l'information. Les nœuds représentent les informations et les arêtes orientées ont, cette fois, pour signification "étend" (fig.5.7). Ainsi nous pouvons lire : la "vitesse du vent" "étend" l'information de "cap".

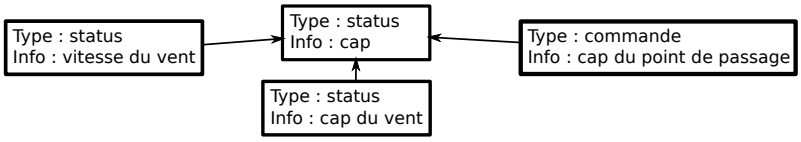

FIG. 5.7 : Exemple d'un graphe d'extension de l'information

Il est ici nécessaire de compléter la frame qui décrit l'information avec suffisamment de terminaux pour pouvoir décrire toutes les relations d'extensions qui existent.

5.3.3 Modèle de dialogue et graphes d'informations

Nous avons présenté le modèle de grounding de Traum ainsi que les ajouts que nous avons proposés. Après avoir abordé une représentation de l'information, intéressons nous maintenant à l'interaction de ces deux modèles aux travers d'un exemple issue du contrôle de drone.

5.3.3.1 Explication de l'information

Un opérateur réalise une mission à l'aide d'un drone. Afin de contrôler ce dernier, l'opérateur dispose d'une interface qui indique la direction, la vitesse et l'altitude du drone. L'opérateur surveille les opérations du système, nous avons donc affaire à la fonction de surveillance du contrôle supervisé. Nous avons adapté le modèle de dialogue de Traum à cette fonction afin d'y observer des mécanismes de contrôle mis en jeu par l'opérateur.

Prenons par exemple l'information "direction". C'est une information de type statut dont le contenu est le cap de l'UAV. Elle est donc présentée soit en continu soit sur demande. En ce qui concerne notre exemple, elle est présentée en continue par le système. Le réseau de transition associé à cette information est donc à l'état 1. Par ailleurs, le nœud du graphe qui correspond à l'information de "direction" pointe sur l'information "ravitailler" comme explication du cap actuelle de l'UAV.

En ce qui concerne les graphes d'explication et d'extension de l'information, l'information initialement pointée est la "direction". Le contenu informationnel à l'état 1 du réseau de transition est lié aux informations pointées dans les graphes d'informations (fig.5.8).

Lorsque l'opérateur souhaite contrôler l'information et la comprendre, il formule une demande d'explication. Le réseau de transition qui traduit l'état du dialogue passe à l'état 2 (fig.5.9). Cet état ne partage aucun lien informatif. C'est un état transitoire qui amène à une modification des informations liées à l'état 1 suite à une réponse du système.

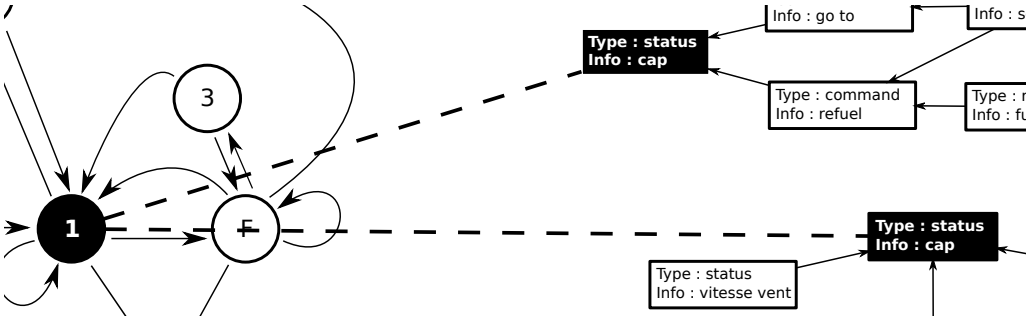

FIG. 5.8 : Interconnexion entre le réseau de transition et les graphes d'explication et d'extension de l'information. A l'état *1*, le RT est lié à l'information introduite initialement.

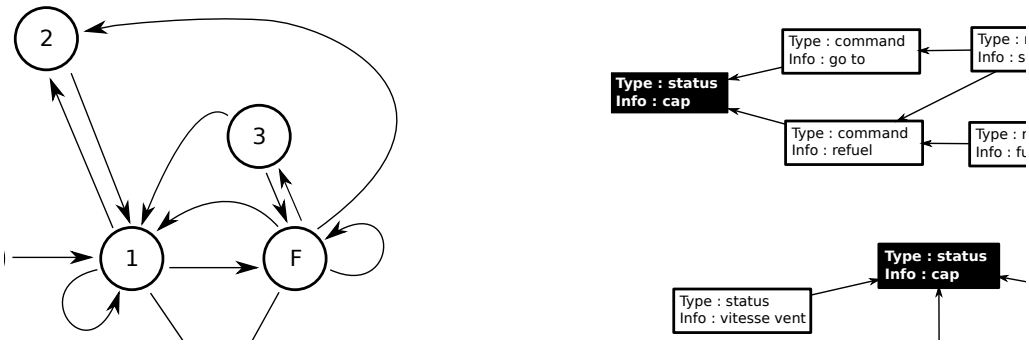

FIG. 5.9 : Interconnexion entre le réseau de transition et les graphes d'explication et d'extension de l'information. A l'état *2*, l'information introduite est toujours active mais n'est pas lié à l'état *2*.

Pour répondre, le système, au sein du graphe, utilise le nœud correspondant à l'information initiale ; celui-ci pointe sur une information qui est l'information

explicative. Dans notre exemple, on se déplace donc au sein du graphe d'explication vers l'information "ravitailler". Lorsqu'elle est communiquée à l'opérateur, le réseau de transition repasse à l'état 1 et pointe maintenant sur cette nouvelle information (fig.5.10).

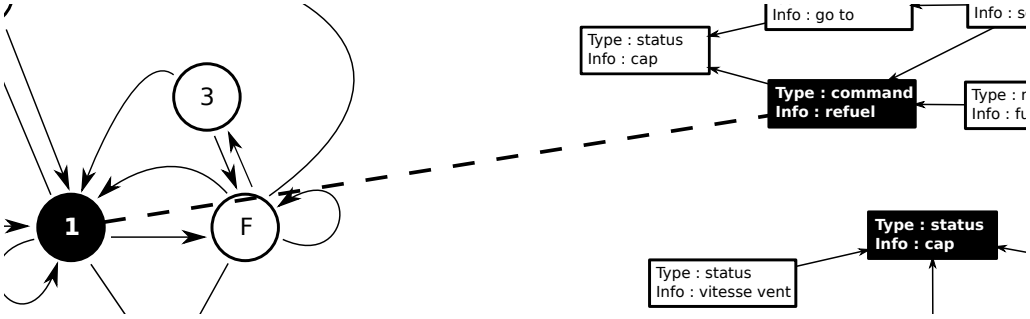

FIG. 5.10 : Interconnexion entre le réseau de transition et les graphes d'explication et d'extension de l'information (A l'état *1* après une demande d'explication à l'information explicative au sein du graphe d'explication).

Il est à rappeler que ce mécanisme exprime une méfiance de l'opérateur qui ressent le besoin de contrôler plus en profondeur le comportement du système. Comme nous l'avons expliqué lors des modifications du modèle de Traum, nous cherchons à observer ces mécanismes pour évaluer la confiance de l'opérateur. On observe que parcourir le graphe d'explication revient à se méfier du système et à remettre en cause la confiance qui a pu lui être donnée.

Enfin un acquittement de l'opérateur clôt le dialogue. Celui-ci est réouvert par le système (monitoring), et l'information pointée sera "direction", c'est-à-dire l'information de départ. On se retrouve dans la configuration initiale de notre exemple.

5.3.3.2 Extension de l'information

Prenons la même situation que précédemment à la différence que cette fois l'opérateur souhaite avoir des informations complémentaires. Au niveau de la modélisation du dialogue, le réseau de transition est à l'état 1 et pointe vers l'information "direction" au sein des graphes d'informations (fig.5.8).

La requête de l'opérateur amène le réseau de transition à l'état 2 (fig.5.9). Cette fois, pour générer la réponse, le système va utiliser le terminal d'extension de l'information de la frame qui représente l'information initiale afin de trouver les informations complémentaires. Les informations pointées au sein du graphe d'extension s'étendent ainsi à ces informations.

La réponse du système ramène le réseau de transition à l'état 1 qui dorénavant pointe sur l'information initiale et les informations complémentaires sélectionnées par le système (fig.5.11).

L'acquittement de l'opérateur clôt le dialogue qui est réouvert par le système pour actualiser l'information de départ. On se retrouve alors dans la configuration

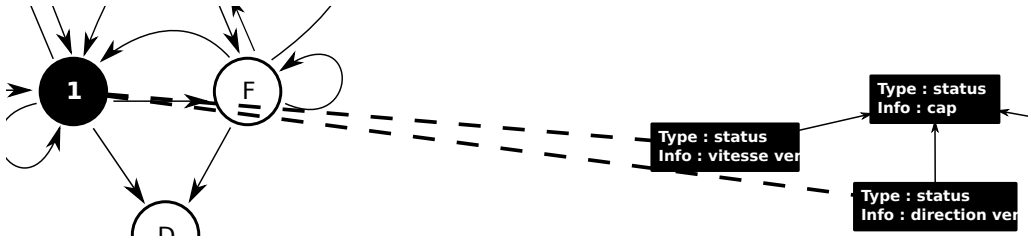

FIG. 5.11 : Interconnexion entre le réseau de transition et les graphes d'explication et d'extension de l'information. A l'état *1* après une demande d'extension de l'information, l'information initiale est toujours pointée ainsi que les informations complémentaires.

initiale (fig.5.8).

Ceci montre le deuxième mécanisme de contrôle adjoint au modèle de Traum. Cette fois-ci, plus le nombre d'informations pointées dans le graphe d'extension de l'information est important, plus l'opérateur se méfie de son système et donc compare un grand nombre d'informations afin de décider de son comportement et de réévaluer sa confiance.

5.4 GESTIONNAIRE DE DIALOGUE POUR L'ÉVALUATION DE LA CONFIANCE

Nous avons présenté les apports au modèle de Traum afin de pouvoir évaluer la confiance d'un opérateur. Notre objectif est de pouvoir faire cette évaluation grâce à l'analyse du comportement d'un opérateur au travers de ce modèle. Voyons comment mettre en forme les données réelles du dialogue au sein de notre modèle pour en extraire les mécanismes de contrôle mis en œuvre par l'opérateur. Nous rappelons que nous travaillons avec l'hypothèse que contrôle et confiance sont deux notions liées par une relation d'opposition dans un cadre coopératif [35].

5.4.1 Gestionnaire de dialogue

Le premier point à aborder est le fonctionnement du gestionnaire de dialogue. Jusqu'à présent nous avons discuté du modèle mais pas de son implémentation à laquelle nous allons nous intéresser maintenant.

Le modèle de Traum repose sur la mise en relation des actes de dialogues et des actes de grounding — qui correspondent aux différentes transitions de son modèle. En effet, l'exécution d'une transition au sein de son modèle est liée à l'interprétation d'un acte de dialogue. Nous définissons ce dernier comme un triplet composé de la manière suivante :

- le locuteur qui réalise l'acte de dialogue (système ou opérateur),
- la forme de l'énoncé que prend l'acte de dialogue (affichage, clic, double clic et focus),

– le contenu sémantique à savoir l'information concernée par l'acte de dialogue.

La forme de l'énoncé et le contenu sémantique forme un acte locutoire. Le gestionnaire de dialogue a pour rôle d'interpréter les actes de dialogue de l'opérateur et de générer les actes de dialogue du système (les réponses).

De plus le gestionnaire assure le suivi des différentes dialogues entre l'opérateur et le système. En effet, dans le cadre de la supervision multi-drones, le système conduit plusieurs dialogues en parallèle pour informer l'opérateur de la localisation de chacun des drones. A chaque dialogue, que l'on nomme unité de dialogue, correspond un réseau de transition, c'est-à-dire une évolution propre à chaque dialogue de l'évolution de l'état de l'information. C'est d'ailleurs sur cette évolution que repose notre évaluation de la confiance puisque nous voulons observer la mise en œuvre des mécanismes de contrôle par l'opérateur.

La première tâche du gestionnaire de contrôle consiste à associer les actes de dialogue qu'il reçoit en entrée à l'unité de dialogue correspondante. Le gestionnaire réalise donc un routage des actes de dialogue (fig.5.12).

Ensuite le gestionnaire de dialogue met à jour les unités de dialogues dont l'information a évolué (typiquement les informations de suivi) et l'unité de dialogue qui s'est vu attribuer l'acte de dialogue en entrée. A partir du nouvel état des unités de dialogue, le gestionnaire génère des actes de dialogue en sortie qui devront être mis en forme par l'interface utilisateur.

Ce processus de traitement du dialogue est illustré par la figure fig.5.12.

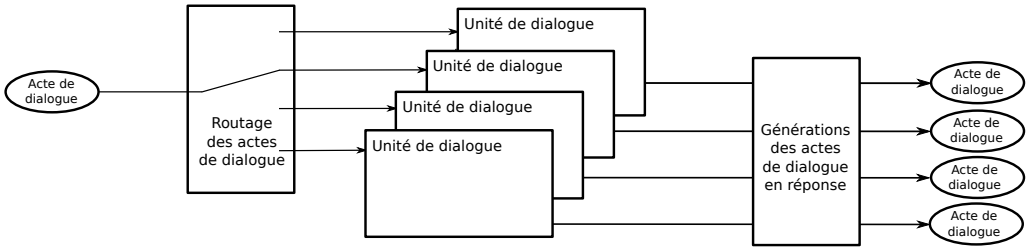

FIG. 5.12 : Gestionnaire de dialogue.

Au sein de ce processus, il nous suffit alors de relever d'une part les interventions de l'opérateur, et d'autre part la lecture qu'il a de l'information.

5.4.2 Suivi du routage des actes de dialogue

Il est à noter qu'un acte de dialogue ne peut pas toujours être associé à une unité de dialogue existante. En effet, l'affichage d'une nouvelle alarme ne peut être attaché à aucune alarme antérieure. Le gestionnaire de dialogue doit alors créer une nouvelle unité de dialogue. Ainsi, la mesure réalisée au niveau du routage des actes de dialogues est l'établissement d'un historique (temporel) des actes de dialogue traités et des actions résultantes caractérisées par :

- le temps : heure à laquelle est traité l'acte de dialogue,
- une action : associer l'acte de dialogue, créer une nouvelle unité de dialogue à partir de l'acte de dialogue ou mettre à jour une unité de dialogue à partir de l'acte de dialogue en entrée,
- un acte de dialogue qui est traité et se décompose selon les trois valeurs présentée précédemment (locuteur, forme de l'énoncé, contenu sémantique).

Prenons un exemple[4]. Un opérateur supervise un ensemble de drones au travers d'une interface graphique qui lui présente l'emplacement des drones sur une carte. Plusieurs drones sont en train d'intercepter des ennemis. L'affichage des interceptions génère la séquence d'actes de dialogues suivante :

1. *système, afficher, interception 1 ;*
2. *système, afficher, interception 2 ;*
3. *système, afficher, interception 1 ;* //les drones se rapprochent de leur position d'interception
4. *système, afficher, interception 2 ;*

Le gestionnaire de dialogue traite les actes de dialogue un à un. Dans un premier temps il va donc essayer d'associer le premier acte de dialogue (système, afficher, interception 1) à une unité de dialogue. Ceci va donc se traduire dans l'historique du routage des actes de dialogue par :

- temps 1 ; associate ; sys → display → interception-1

Actuellement il n'y a aucune unité de dialogue, le gestionnaire doit donc en créer une pour l'information drone 1 :

- temps 1 ; create ; 0 ; sys → display → interception-1

Le même processus est appliqué au deuxième acte de dialogue. On obtient alors dans l'historique les lignes suivantes :

- temps 2 ; associate ; sys → display → interception-2
- temps 2 ; create ; 1 ; sys → display → interception-2

Lors du troisième acte de dialogue, nous avons deux unités de dialogues actives. Or ce nouvel acte de dialogue concerne l'interception 1, nous allons donc avoir cette fois une mise à jour de l'unité de dialogue associée. Nous avons alors les lignes suivantes (illustrées figure fig.5.13) :

- temps 3 ; associate ; sys → display → interception-1
- temps 3 ; update ;0 ; sys → display → interception-1

Enfin le dernière acte de dialogue est traité de la même manière mais il sera associé à l'unité de dialogue qui concerne l'interception 2.

Ici, l'élément important à mesurer est le type de dialogue qui est initialisé. En effet, au chapitre 4, on a observé — qualitativement — que le taux d'intervention d'un opérateur était corrélé à la confiance de l'opérateur. L'analyse du routage des actes de dialogue est une manière simple de mesurer au travers du dialogue ce taux d'intervention.

[4]issue de l'expérimentation présentée au chapitre suivant

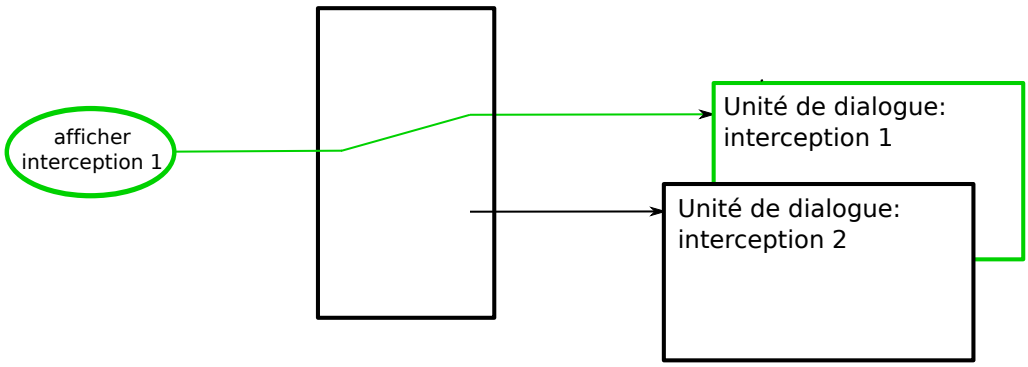

FIG. 5.13 : Association d'un acte de dialogue à une unité de dialogue existante.

5.4.3 Suivi de la dynamique des unités de dialogue

Bien entendu, l'historique du routage des actes de dialogues n'est pas suffisant en soit. Nous avons vu précédemment que l'évaluation de la confiance repose sur une quantification des mécanismes de contrôle mis en œuvre par un opérateur. Pour cela, il nous faut connaître l'évolution de chacune des unités de dialogue.

Pour ce faire, les informations pertinentes à retenir lors du dialogue sont :

- le temps : heure à laquelle l'unité de dialogue subit un changement d'état,
- l'information concernée par l'unité de dialogue,
- le type de dialogue, s'il se rapporte à du monitoring ou à une intervention,
- l'état courant du dialogue,
- le locuteur,
- l'acte grounding déclenché par l'acte de dialogue,
- le nouvel état du dialogue.

Si nous conservons ces informations, nous pouvons donc transposer l'interaction dans le formalisme de notre modèle de dialogue. Prenons, la séquence d'acte de dialogue suivante pour illustrer ce propos :

1. *système, afficher, position du drone 2 ;*
2. *système, afficher, position du drone 2 ;*
3. *opérateur, focus, drone 2 ;*
4. *système, afficher, position du drone 2 ;*
5. *opérateur, double clique, drone 2 ;*
6. *système, pop-up, drone 2 ;*

La première tâche du gestionnaire de dialogue est d'associer ces actes de dialogue à l'unité de dialogue approprié. Pour cet exemple, l'unité de dialogue est celle concernant le drone 2. Une fois l'acte de dialogue associé à une unité de dialogue, le gestionnaire de dialogue met à jour l'état de l'unité de dialogue et génère un acte

de dialogue en réponse si nécessaire. Celui-ci sera traité, par l'interface utilisateur, et mis en entrée du gestionnaire de dialogue.

Si nous prenons le premier acte de dialogue, celui-ci va initier une nouvelle unité de dialogue comme nous l'avons vu dans la section précédente. Le dialogue lors de son initialisation est à l'état 0, soit l'état initial. La transition activée est la transition 'initiate()'. Ainsi l'historique, à propos de l'unité de dialogue de l'information drone 2, débutera avec une entrée du type :

– temps 1 ; drone 2 ; monitoring ; 0 ; système ; initiate ; 1

Le deuxième acte de dialogue est généré par le système afin de communiquer une mise à jour de la position du drone. Cet acte de dialogue se traduit alors par une transition repair, avec pour entrée dans l'historique la ligne suivante :

– temps 2 ; drone 2 ; monitoring ; 1 ; système ; repair ; 1

Le focus attentionnel de l'opérateur fait partie des actes de dialogue disponibles pour interagir avec le système. Il se traduit dans notre modèle par un acquittement de l'information concernée. Nous avons donc dans l'historique pour cet acte de dialogue la ligne suivante :

– temps 3 ; drone 2 ; monitoring ; 1 ; opérateur ; ack ; 5

L'état du dialogue passe à l'état 5, état final du dialogue. Or nous sommes dans une unité de dialogue de type monitoring, c'est à dire une unité de dialogue où l'information est constamment mise à jour. L'unité de dialogue sera donc ouverte à nouveau et ainsi ramenée à l'état 1. Ainsi, lors du prochain rafraîchissement de l'information, le système génère l'acte de dialogue numéro 4 qui se traduit dans l'historique par la ligne :

– temps 4 ; drone 2 ; monitoring ; 5 ; système ; repair ; 1

En ce qui concerne un contrôle de l'information, l'acte de dialogue suivant permet à l'opérateur de consulter des informations complémentaires à propos du drone 2. Ce mécanisme de contrôle a été présenté précédemment dans ce chapitre. Cette fois le gestionnaire de dialogue a une troisième tâche à accomplir : la génération d'une réponse appropriée. Dans notre exemple, le dernier acte de dialogue est généré par le système comme réponse à l'opérateur. A charge pour l'interface de le formuler. Lorsque l'acte de dialogue est formulé par l'interface, celui-ci est mis en entrée du gestionnaire de dialogue pour être traité par celui-ci. Nous avons alors la séquence suivante dans l'historique de l'unité de dialogue :

– temps 5 ; drone 2 ; monitoring ; 1 ; opérateur ; reqext ; 2
– temps 6 ; drone 2 ; monitoring ; 2 ; système ; extend ; 1

Cet exemple peut être synthétisé au sein d'un graphe, illustré figure fig.5.14, où les noms des transitions sont remplacés par les actes de dialogues. Ainsi on peut voir que, si à chaque acte de dialogue, dans un contexte donné, est associé un acte de grounding clairement identifié, il est possible de transposer les interactions d'un opérateur et de son système au sein de notre modèle. Cette transposition doit nous permettre, comme nous allons le voir dans la section suivante 5.5, d'analyser l'évolution du dialogue et ainsi d'évaluer le degré de confiance de l'opérateur — à partir d'une quantification des mécanismes de contrôle mis en jeu.

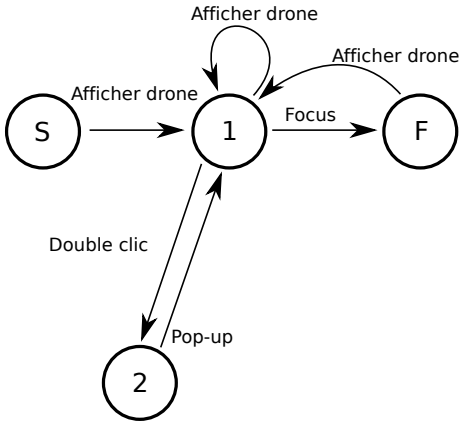

FIG. 5.14 : Association d'un acte de dialogue aux transitions du modèle de dialogue.

5.5 MODÉLISATION DE L'ÉVALUATION DE LA CONFIANCE

Nous proposons ici un modèle d'évaluation de la confiance basé sur l'analyse de notre modèle de dialogue. Les différentes hypothèses, introduites ici, seront validées au chapitre suivant.

Opposition entre degré de confiance et volume d'information : la confiance $Trust$ est en opposition au volume d'informations échangées, c'est-à-dire au nombre d'unité de dialogue NUD. On a alors $Trust = -\alpha NUD$, α est une constante strictement positive.

Opposition entre degré de confiance et nombre d'intervention : la confiance $Trust$ est en opposition au volume d'interventions de l'opérateur $NUDIntervene$. On a alors $Trust = -\alpha NUDIntervene$, α est une constante strictement positive.

Opposition entre degré de confiance et focus attentionel : la confiance $Trust$ est en opposition au nombre d'acquittement de l'opérateur aux informations par focus attentionel $NAck$. On a alors $Trust = -\alpha NAck$, α est une constante strictement positive.

Opposition entre degré de confiance et demande d'explication : la confiance $Trust$ est en opposition au nombre de demandes d'explication $NReqExpl$. On a alors $Trust = -\alpha NReqExpl$, α est une constante strictement positive.

Opposition entre degré de confiance et demande d'information étendue : la confiance $Trust$ est en opposition au nombre de demandes d'information étendue $NReqExt$. On a alors $Trust = -\alpha NReqExt$, α est une constante strictement positive.

6 Validation expérimentale du modèle

Afin de valider notre modèle une campagne expérimentale a été mise au point. Celle-ci nous permet d'affiner notre analyse du dialogue par rapport à la précédente campagne expérimentale (présentée au chapitre 4). Dans un premier temps nous compléterons la description de la plateforme expérimentale à l'aide du gestionnaire de dialogue introduit au chapitre précédent. Puis nous présenterons le protocole expérimental et les résultats obtenus.

6.1 Evolution du simulateur de contrôle multi-drones

6.1.1 Nouvelle granularité des tâches

Nous avons vu lors d'une expérimentation préliminaire (chapitre 4) que le séquençage de la tâche d'interception en agrégation prédiction et interception semblait être réalisé mentalement par l'opérateur. C'est pourquoi nous avons décidé dans cette campagne expérimentale de ne plus distinguer ces tâches. Bien entendu, le système suit toujours cette procédure. Nous n'avons donc plus que trois tâches :

- la détection
- l'interception
- le ravitaillement

L'automatisation des tâches n'est plus configurable. Nous faisons ce choix pour deux raisons :

- l'évaluation de la confiance repose sur l'interaction entre l'opérateur et l'automate, si ce dernier est éteint il n'y a plus d'interaction possible. Nous ne pouvons plus évaluer la confiance même si nous pouvons supposer une forte méfiance de l'opérateur au vu du rejet de l'automate. Ce cas extrême n'est donc pas important pour notre étude.
- lors de la première expérience, l'extinction de l'automate a été constatée pour les fonctions d'agrégation et de prédiction (22% des sujets), or ces fonctions ne sont plus accessibles à l'opérateur. La possibilité de configurer l'automate devient inutile au vu de cette première expérience. Et comme nous l'avons dit plus haut l'extinction de l'automate est un cas extrême d'expression de la méfiance qui n'est pas utile à notre approche.

6.1.2 Simplification de l'interface

Nous venons de voir que l'activation ou la désactivation d'un automate n'est pas utile à notre étude. Nous avons donc décidé de simplifier le panel d'actions. Ainsi, il ne se compose plus que de quatre boutons qui correspondent aux différentes commandes accessibles à l'opérateur et de deux boutons pour la validation et la remise à zéro[1] d'une commande. Ce panel est complété d'un texte qui résume l'état de la commande que l'opérateur est en train d'élaborer (fig.6.1).

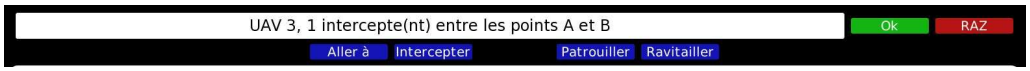

FIG. 6.1 : Panel de commande de la nouvelle interface graphique.

La seconde modification concerne l'accès aux informations, nous voulons minimiser les informations directement accessibles par l'opérateur. En effet, si l'opérateur souhaite vérifier des niveaux de fuels, c'est qu'il remet en cause le fonctionnement de l'automate (fiabilité et/ou stratégie). L'évaluation de la confiance repose sur la mesure des comportements de dialogue qui soient représentatifs de mécanismes de contrôle (telle qu'expliquée au chapitre 5). Par exemple, l'accès au fuel d'un UAV est une demande de complément d'information concernant l'UAV. Les informations complémentaires ou les explications sont disponibles à l'aide de pop-up qui s'ouvre à l'aide d'un double clic sur la représentation graphique d'une information initiale (fig.6.2).

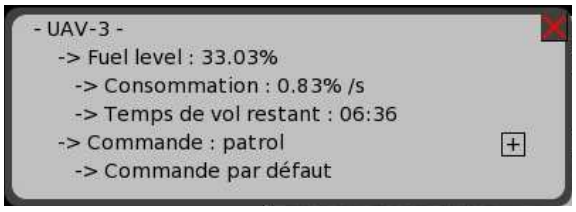

FIG. 6.2 : Pop-up informatif pour l'obtention d'informations complémentaires et d'explications.

Enfin le panel tactique reste identique à la version utilisée lors de l'expérience préliminaire (chapitre 4).

6.1.3 Gestionnaire de dialogue

Nous avons vu au chapitre précédent comment nous modélisons le dialogue ainsi que le fonctionnement global du gestionnaire de dialogue. Intéressons nous à présent

[1] tant que la commande n'est pas validée

à la relation entre l'interface utilisateur et le gestionnaire, c'est-à-dire l'interprétation des actes de dialogue émis par le système.

6.1.3.1 Information de type suivi

Les informations de type monitoring sont :
- la localisation des drones,
- la localisation des alarmes,
- la localisation des interceptions.

Leur durée de vie n'est pas la même selon leur type. Ainsi une interception sera affichée le temps qu'elle dure, une alarme un temps limité (2 minutes) et les drones durant toute la simulation sauf s'ils disparaissent faute de fuel. Mais dans tout les cas le schéma de dialogue est le même. Prenons le cas d'une interception pour illustrer le lien entre l'interface utilisateur et le gestionnaire de dialogue.

Lorsqu'une interception est mise en œuvre par le système, celle-ci est affichée (fig.6.3) et est régulièrement mise à jour. L'opérateur peut prendre connaissance de l'information en focalisant son attention visuelle sur celle-ci. Enfin lorsque l'interception est réussi et qu'un laps de temps minimum s'est écoulé, l'information d'interception disparaît de l'affichage.

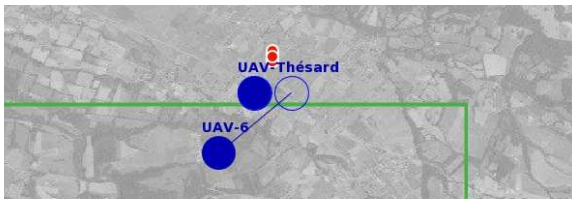

FIG. 6.3 : Affichage d'une interception.

Dans le gestionnaire de dialogue, ceci se traduit par la création d'une nouvelle unité de dialogue, sa mise à jour et enfin son abandon. Par ailleurs, le focus visuel de l'opérateur se traduit par un acquittement au sein de l'unité de dialogue. Nous obtenons ainsi au sein de notre modèle la séquence illustrée fig.6.4.

L'acquittement visuel d'une information est un mécanisme basique de contrôle de l'information. Nous allons maintenant illustrer comment sont mis en œuvre les mécanismes d'explications et d'extensions de l'information. Rappelons que notre modèle de dialogue met en avant ces mécanismes car nous supposons qu'ils sont des indicateurs de confiance. Il est donc important que notre interface permette à l'opérateur d'entreprendre de telles actions.

6.1.3.2 Contrôle de l'information

Le premier mécanisme de contrôle est l'acquittement — qui existe de base dans le modèle de Traum — qui se concrétise, avec notre plateforme, par un focus attentionel

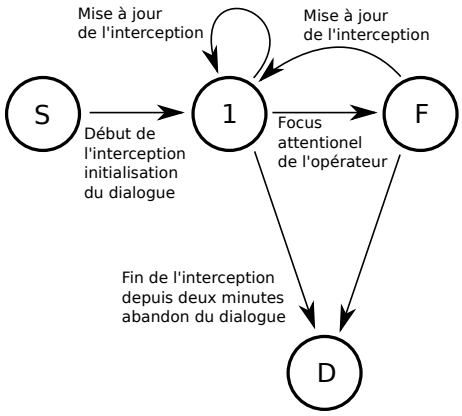

FIG. 6.4 : Illustration de l'activité de suivi.

. Afin de pouvoir mesurer, un eye-tracker a été utilisé. L'interface récupère au travers d'une API les coordonnées du focus attentionnel sur l'écran. Elle peut ainsi vérifier si le regard de l'opérateur se pose ou non sur un objet informatif (affichage des drones, des alarmes, etc.). Lorsque cela arrive un acquittement est généré et transmis au gestionnaire de dialogue. Dans celui-ci, l'unité de dialogue concernée passe à l'état 'F' (transition *ack()*).

Nous avons aussi modifié le modèle de Traum afin d'inclure des mécanismes d'extension de l'information et d'explication. Pour cela nous avons inclus dans notre interface un système de pop-up qui s'ouvre lors d'un double clic de l'opérateur sur certains objets graphiques (alarmes, drones, visualisations de commande e.g. interception).

Si nous reprenons notre exemple précédent, l'opérateur peut dans un premier temps obtenir des compléments d'information sur l'interception à l'aide d'un double clic qui ouvre une pop-up (fig.6.5). Ces nouvelles informations peuvent à leur tour être sujet à une demande d'information complémentaire (icône '+') ou une demande d'explication (icône ' ?').

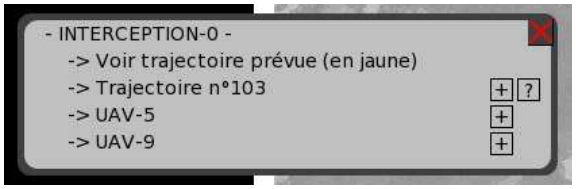

FIG. 6.5 : Pop-up d'information d'une interception.

Au sein du gestionnaire de dialogue, l'unité de dialogue qui est associée au suivi de cette interception va passer à l'état '2' chaque fois que l'opérateur effectue une demande d'extension de l'information ou une demande d'explication. Elle retourne

à l'état '1' à l'affichage de la réponse. Quant à la fermeture de la pop-up par l'opérateur, elle génère un acquittement et amène l'unité de dialogue à l'état 'F'. Comme on a pu le voir précédemment dès la prochaine mise à jour de l'interception, l'unité de dialogue repasse à l'état '1'. Au niveau de notre modèle, l'interaction avec notre interface est illustrée, pour cette séquence, par la figure 6.6.

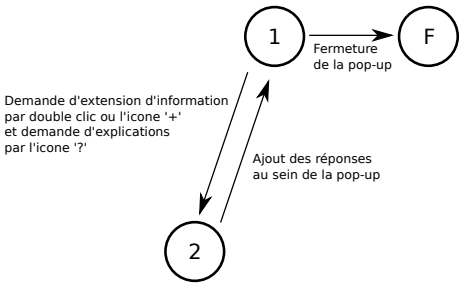

FIG. 6.6 : Illustration des demandes d'extension de l'information et des demandes d'explication.

6.1.3.3 Elaboration de commande

Dans le cadre de l'élaboration d'une commande, on distingue trois catégories de dialogue :

- la sélection/désélection des drones, et des commandes ;
- la validation d'une commande ;
- la réinitialisation d'une commande.

Elles suivent toutes un même processus : l'opérateur clique sur un élément graphique et le système acquitte par un retour visuel (e.g. surbrillance des drones sélectionnés). De plus au cours de l'élaboration d'une commande et jusqu'au dialogue de validation (clic du bouton "ok"), la pré-visualisation textuelle de la commande est mise à jour (fig.6.1).

Contrairement à notre modèle présenté au chapitre 5, les mécanismes de contrôle pour les dialogues d'élaboration de commandes ne sont pas implémentés. En ce qui concerne les commandes du système, ce dernier a un niveau 9 d'autonomie sur l'échelle de Sheridan. Il est autonome et ne fait qu'informer l'opérateur de son activité. L'opérateur ne peut donc pas intervenir lors de l'élaboration de la commande. Il ne pourra que donner des contre-ordres pour corriger l'activité du système. Pour ce qui est des commandes initiées par l'opérateur, aucune suggestion n'est fournie par le système.

6.1.4 Générateur de scénarios

Pour cette campagne expérimentale, nous avons créé un générateur de scénarios aléatoires. Le scénario résultant respecte un certain nombre de contraintes :

– Les scénarios sont d'une durée de dix minutes ;
– le point d'entrée sur la carte des intrus est aléatoire le long du bord de la carte ;
– un nouvel intrus apparaît toutes les trente secondes ;
– la cible de l'intrus est aléatoire ;
– la trajectoire pour atteindre sa cible est semi-rectiligne.

Nous définissons une trajectoire semi-rectiligne comme une trajectoire rectiligne qui a été segmentée et dont les points de jonctions entre segments sont écartés de leur position aléatoirement. Une telle trajectoire est illustrée figure 6.7.

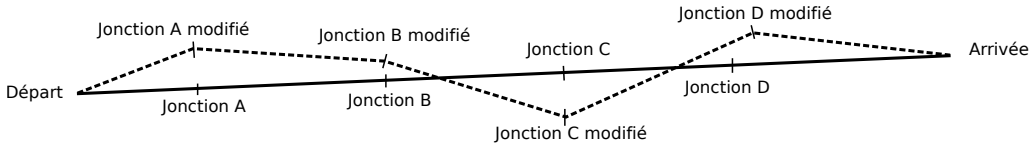

FIG. 6.7 : Illustration d'une trajectoire semi-rectiligne (en pointillé).

6

6.2 PROTOCOLE EXPÉRIMENTAL

6.2.1 Mesure

6.2.1.1 Quantitative : stratégie d'interaction

Nous souhaitons vérifier l'hypothèse que la confiance est exprimée au cours du dialogue au travers du comportement. Pour cela nous avons relevé au sein de la littérature le lien d'opposition qui existe entre confiance et contrôle dans le cadre de la coopération. C'est pourquoi nous proposons ici des mesures objectives du comportements de l'opérateur. Ainsi sont mesurées les actions élémentaires suivantes :

– le focus visuel : lorsque le regard de l'opérateur s'attarde sur un objet graphique ;
– la sélection/désélection : lorsque l'opérateur clique sur différents éléments pour élaborer une commande ;
– la validation d'une commande : lorsque l'opérateur après avoir défini les paramètres d'une commande clic sur le bouton de validation ;
– le contrôle de l'information : lorsque l'opérateur accède à des informations complémentaires (double clic sur un élément informatif e.g. uav) et parcours le contenu des pop-up résultantes (requête supplémentaire d'informations complémentaires et d'explications).

De plus nous enregistrons l'évolution de chacune des unités de dialogue qui auront eu lieu au cours de la session. Ainsi nous aurons, pour une information donnée, l'ensemble des transitions générées par les actions élémentaires de l'opérateur.

6.2.1.2 Qualitative : confiance

La confiance est une donnée subjective. Nous l'évaluons à l'aide du questionnaire de Jian [54] qui a toutefois l'inconvénient de ne donner qu'une évaluation globale à l'ensemble d'une session. Or plus l'expérience est longue, plus la stratégie d'interaction du sujet à des chances d'évoluer au sein de la session. C'est pourquoi il nous faut réaliser une évaluation de la confiance à une échelle de temps plus réduite. Pour cela, nous proposons une deuxième méthode d'évaluation qui consiste à visionner la session en posant les questions suivantes au sujet :

question sur évènement d'affichage Lors de l'évènement X, quel était votre degré de confiance/méfiance concernant l'intervention du système, sur une échelle de 1 à 5 ? (1 très méfiant - rejet du système, 5 très confiant - aucun doute). A ces évènements correspondent l'apparition des commandes de ravitaillement et d'interception.

question sur une action de l'opérateur Vous avez fait l'action X, remettez vous en cause le fonctionnement de l'automate ? Si oui, à quel degré évalueriez vous votre confiance/méfiance, sur une échelle de 1 à 5 ?

généralisation à partir d'un évènement Cela a-t-il modifié votre ressenti vis-à-vis du système dans son ensemble ? (1 très méfiant - rejet du système, 5 très confiant - aucun doute).

Afin d'anticiper dans l'explication d'écart comportementaux entre les sujets, il leur sera demandé de donner une définition succincte de la confiance en l'automatisme.

6.2.2 Déroulement de l'expérimentation

Au final nous proposons le déroulement suivant pour un sujet.

6.2.2.1 Introduction et apprentissage

L'expérimentation est présentée comme une étude sur l'analyse de l'interaction dans le cadre de la supervision multi-UAVs. Un scénario très simple (arrivée d'intrus par 4 directions différentes sur des trajectoires rectilignes en boucle) permet la prise en main du système et l'apprentissage des commandes de ce derniers. Cette prise en main dure 15 minutes.

Le questionnaire de panel est rempli par le sujet au cours de cette étape.

6.2.2.2 Session 1

Scénario : la première session est caractérisée par l'arrivée d'intrus solitaires. Leur point d'entrée est aléatoire, leur cible de destination est aléatoire et leur trajectoire semi-rectilignes. C'est le scénario n°1.

La durée du scénario est d'environ 10 minutes.

Ensuite une définition de la confiance est demandée au sujet.

La séance se poursuit avec le questionnaire de Jian et puis un visionnage pour l'évaluation de la confiance. L'objectif de l'expérimentation est ainsi dévoilé au sujet.

Objectif : la première session a pour but de constituer d'un panel témoin. En effet avec un panel de 24 personnes, nous ne pouvons pas le diviser en deux pour constituer un groupe témoin. L'idée est donc d'utiliser les données de cette première session en tant que données issues d'un groupe témoin.

6.2.2.3 Session 2

Scénario : le scénario n°1 est utilisé pour cette session (scénario de la session 1).

La durée du scénario est d'environ 10 minutes.

La séance se poursuit avec le remplissage du questionnaire de Jian ainsi qu'un visionnage pour l'évaluation de la confiance.

Objectif : cette seconde session va permettre d'évaluer l'impact du biais introduit par la connaissance de l'objectif expérimental par le sujet, s'il est observable (non négligeable). On devrait a priori y constater une accentuation des comportements liés à la confiance due à l'attention que le sujet va porter à sa confiance.

6.2.2.4 Session 3

Scénario : le scénario n°1 est utilisé pour cette session (scénario de la session 1). Par contre les capacités du système seront dégradées : le calcul de positionnement des interceptions sera bruité afin de provoquer des échecs[2]. Les drones sont ainsi écartés de leur position d'interception rendant celle-ci inefficace.

La durée du scénario est d'environ 10 minutes.

La séance se poursuit avec le remplissage du questionnaire de Jian ainsi qu'un visionnage pour l'évaluation de la confiance.

Objectif : le dysfonctionnement du système pour le calcul d'interception devrait provoquer une méfiance du sujet vis-à-vis du système. C'est cette dégradation de la confiance que nous souhaitons observer à l'aide de cette session.

6.2.2.5 Session 4

Scénario : le scénario n°1 est utilisé pour cette session (scénario de la session 1).

[2]Les coordonnées de positionnement de chaque drone sont calculées normalement puis à chaque coordonnées est ajoutées une valeur aléatoire (prise entre -15 et 15 en ce qui nous concerne). L'amplitude de valeur choisit influence directement le taux d'échec

La durée du scénario est d'environ 10 minutes.

La séance se poursuit avec le remplissage du questionnaire de Jian ainsi qu'un visionnage pour l'évaluation de la confiance.

Objectif : ici nous essayons de voir quels sont les effets résiduels de la session 3 en revenant à un scénario pour lequel l'automate est capable. On cherche à observer la reconstruction de la confiance. Et si elle a lieu de voir si le comportement obtenu est proche ou similaire à celui de la session 2.

Nous obtenons la différenciation suivante sur les quatre sessions expérimentales :
- session 1 : pas de connaissance a priori sur les évaluations post-session, automates parfaitement fonctionnels,
- session 2 : connaissance sur les évaluations post-session, automates parfaitement fonctionnels,
- session 3 : connaissance sur les évaluations post-session, dysfonctionnement du système d'interception,
- session 4 : connaissance sur les évaluations post-session, automates parfaitement fonctionnels.

6.3 Panel

La population est fortement issue du milieu scientifique et technique. En effet, la plupart des sujets ont été recrutés parmi les élèves de Télécom Bretagne. On peut donc supposer que la population, si elle n'est habituée à manipuler des drones, a d'importantes affinités avec les systèmes informatiques actuels.

La répartition Homme-femme est :
- Homme : 91,7%
- femme : 8,3%

La parité Homme-femme est fortement déséquilibré. En effet, la population étudiante de Télécom Bretagne est fortement dissymétrique. Il est donc normal que l'on retrouve cette disparité dans notre groupe de sujets.

6.4 Etude de la confiance : analyse inter-session

Avant de valider notre modèle, nous avons trois points à contrôler vis-à-vis de notre protocole expérimental. En effet, nous souhaitons savoir dans un premier temps si la connaissance *a priori* d'une évaluation de la confiance a un effet sur les résultats obtenus. Pour cela nous aurons à comparer les mesures subjectives obtenues avec les questionnaires de Jian des deux premières sessions. Ensuite le second point est de vérifier que la troisième session de notre protocole engendre une dégradation de la confiance de sujet. Enfin nous nous intéresserons aussi à l'observation du rétablissement de la confiance lors de la quatrième session pour voir si la méfiance développée au cours de la troisième session a un effet résiduel lors d'un retour à la normale.

Méthode :

Cette étude inter-session est basée sur les questionnaires de Jian. Nous avons vu au chapitre 2 qu'un certain nombre d'études utilisent les résultats de ce questionnaire à l'aide d'une somme pondérée pour définir un indicateur de confiance. Nous utiliserons cette même démarche en appliquant la formule suivante :

$$Trust_{(i,j)} = \sum_{k=1}^{k=12} \pm Q_{(i,j,k)} \tag{6.1}$$

$Q_{(i,j,k)}$ est la réponse à la question numéro $k \in [\![1;12]\!]$ du sujet i à la session j. Celle-ci contribue positivement '+' ou négativement '−' à la somme selon que la question est liée à la confiance (question 6 à 12) ou à la méfiance (question 1 à 5).

Afin de tester l'existence d'une variation significative nous souhaitons utiliser un test ANOVA. Ce dernier nécessite que la variable $Trust_{(i,j)}$ suit une loi normale. Pour vérifier cela nous appliquerons le test de Shapiro-Wilk.

Enfin, pour comparer les sessions deux à deux, nous utiliserons un test t deux à deux avec "paired" sur la variable $\text{TrustMean}_{(j)}$:

$$TrustMean_{(j)} = \frac{\sum_{i=1}^{i=24} Trust_{(i,j)}}{24} \tag{6.2}$$

Résultats :

Le test de Shapiro-Wilk donne les résultats suivants :

Session	p-value
1	0,33
2	0,42
3	0,6
4	0,26

Avec un seuil de p-value à 5%, nous pouvons conclure que la variable $Trust_{(i,j)}$ suit une loi normale. Il est donc possible d'appliquer un test ANOVA sur celle-ci. Nous obtenons une p-value de $7.768e-08$. Il y a donc des différences significatives entre les moyennes (p-value $< 10^{-7}$). Par contre l'ANOVA ne nous dit pas entre quelles sessions se situent ces différences. Nous avons donc une ou des sessions qui se différencient des autres.

Le test t appliqué à la variable $TrustMean_{(j)}$ donne le résultat suivant :

Session	1	2	3
1	-	-	-
2	0,14	-	-
3	9,30E-006	4,90E-007	-
4	0,4	0,83	1,50E-006

On constate alors qu'il n'y pas de différence significative entre les 1, 2 et 4. La session 3 est significativement différentes des autres (p-value $< 10^{-4}$). La différence se situe lors de la défaillance du système d'interception qui induit une diminution de la confiance dans les évaluations a posteriori. Cette variation est significative, nous en concluons donc en accord avec la littérature que la défaillance du système a un impact sur le degré de confiance de l'opérateur. En effet, d'après le modèle descriptif de la confiance de Lee, l'historique des interactions a un effet sur le niveau de confiance. Or il est tout a fait normal qu'un comportement contre performant du système produise une diminution de la confiance chez l'opérateur. Le deuxième observable est la non-significativité de variation entre les sessions 1 et 2. Cela signifierait en effet que la connaissance d'une évaluation sur la confiance post-session n'a pas d'incidence notable sur les résultats de l'évaluation. Enfin la variation non significative entre les sessions 2 et 4 montre qu'il n'y a pas d'effet notable de la troisième session sur la confiance lors du retour à la normale avec la quatrième session. Attention, ce dernier résultat est à relativiser. En effet, nous étudions une évaluation globale à la session, il peut donc y avoir une variation plus significative au début de la quatrième session mais négligeable sur la durée de la session. Nous approfondirons ce point à la section 6.7.2.

> La troisième session expérimentale est représentative d'une situation de méfiance contrairement aux autres sessions.

6.5 ETUDE DU LIEN ENTRE CONFIANCE ET NOMBRE D'UNITÉ DE DIALOGUE

Nous voulons baser l'évaluation de la confiance sur une analyse du dialogue. Pour cela nous avons défini un modèle qui permet de suivre l'évolution de chacune des informations qui est échangée entre un opérateur et son système. Notre évaluation se base sur la relation d'opposition entre confiance et contrôle. Or le contrôle c'est, selon la définition de Castelfranchi (donnée à la section 5.0.1), la vérification du bon déroulement des tâches ainsi que les interventions nécessaires pour assurer ce bon déroulement. Ainsi la première observation que l'on devrait constater au sein de notre expérimentation est une augmentation de la circulation des informations entre un opérateur et un système. En effet, si l'opérateur devient méfiant, on devrait pouvoir

constater une surveillance du système plus conséquente et un nombre d'intervention en net augmentation. Nous proposons donc dans cette étude de mettre en relation l'évaluation de la confiance et le nombre d'unités de dialogue mises en œuvre durant l'ensemble de la session. Autrement dit, on étudie le lien entre confiance et quantité d'informations.

Méthode :

Afin de vérifier cette hypothèse, nous utilisons la variable $NUD_{(i,j)}$ qui représente le nombre d'unité de dialogue au cours de la session j du sujet i. Elle est obtenue en comptabilisant le nombre d'entrée du type "nouveau" au sein du fichier log de l'appariement des actes de dialogues correspondant à la session j du sujet i.

Pour pouvoir appliquer un test ANOVA sur la variable :

$$NUDMean_{(j)} = \frac{\sum_{i=1}^{i=24} NUD_{(i,j)}}{24} \tag{6.3}$$

Il nous faut au préalable vérifier que la variable $NUD_{(i,j)}$ suit une loi normale avec un test de Shapiro-Wilk. Le test ANOVA nous permet alors de valider l'existence d'une variation significative entre les sessions. Ce résultat est à compléter avec une analyse deux à deux de la variable $NUDMean_{(j)}$ qui permettra d'identifier quelle est la (ou les) session(s) qui est (sont) significativement différente(s). Pour cela nous utiliserons un test t avec "paired".

Résultats :

Le test de Shapirov-Wilk sur la variable $NUD_{(i,j)}$ donne les résultats suivants :

Session	p-value
1	0,053
2	0,32
3	0,87
4	0,19

Ainsi avec un seuil de p-value à 5%, on peut considérer que cette variable suit une loi normale. On peut alors appliquer le test ANOVA pour lequel on obtient une p-value de 4.765e-9. On en conclut donc qu'il y a des différences significatives entre les moyennes (p-value $< 10^{-8}$).

Enfin, on obtient les résultats suivants en appliquant le test t :

Session	1	2	3
1	-	-	-
2	1,02E-002	-	-
3	1,10E-003	4,60E-009	-
4	9,94E-002	6,76E-002	3,70E-009

Nous avons une variation significative de la session 3 par rapport aux autres (p-value $< 10^{-3}$). Ainsi, la défaillance de l'automate a une incidence sur le nombre d'unité de dialogue entre l'opérateur et le système. Il est tout à fait logique qu'en cas de dysfonctionnement du système, un plus grand nombre de dialogue soit mis en œuvre par son opérateur. En effet, ce dernier doit intervenir, ce qui implique l'augmentation des commandes réalisées manuellement. L'opérateur se substitue au système afin de compenser la défaillance de ce dernier. Nous analyserons plus loin quels types de dialogue sont réellement impactés sachant que seul le système d'interception est défectueux.

En effet, ici nous prenons en compte l'ensemble des unités de dialogue que ce soit au niveau de la couche de suivi ou bien au niveau de la couche de programmation/intervention de notre modèle de dialogue. Or l'augmentation des unités de dialogue de la couche de suivi n'est pas liée à l'activité de l'opérateur. En effet, si elle augmente ce n'est pas à son initiative. L'incapacité du système à gérer les interceptions laisse plus de temps aux intrus pour se déplacer ce qui peut induire un plus grand nombre d'alarmes liées à un même intrus. Pour valider l'hypothèse d'un lien entre la quantité d'informations échangée et la confiance, nous ne devons prendre en compte que les informations qui circulent de l'opérateur vers le système. C'est-à-dire prendre en compte uniquement les unités de dialogue de la couche programmation/intervention.

Enfin cette variation a lieu lors de la troisième session, session durant laquelle les sujets ont montré une perte significative de confiance vis-à-vis du système. Regardons alors le lien entre confiance et nombre d'unité de dialogue mais uniquement dans le cadre de la couche d'intervention de notre modèle.

6.5.1 Unité de dialogue de type intervention

Nous avons vu lors de l'expérience préliminaire qu'un comportement de méfiance induisait des interventions plus nombreuses de la part de l'opérateur. Nous souhaitons confirmer ce résultat en vérifiant l'existence d'une corrélation entre le degré de confiance de l'opérateur et le nombre d'unité de dialogue de type intervention.

Méthode :

Pour cette étude nous définissons une nouvelle variable : $NUDIntervene_{(i,j)}$. Elle représente le nombre d'unités de dialogue de type intervention (initiée par l'opéra-

teur). Ce nombre s'obtient en comptabilisant les unités de dialogue qui portent sur l'information "bouton ok" (validation d'une commande).

Afin de vérifier la pertinence de notre hypothèse de travail, nous allons dans un premier temps vérifier l'existence d'une variation significative de cette variable lors de la troisième session. Pour cela nous commencerons par vérifier la distribution de $NUDIntervene_{(i,j)}$ à l'aide d'un test de Shapiro-Wilk. En effet, cela nous permettra de savoir si nous pouvons appliquer un test ANOVA sur la variable :

$$NUDInterveneMean_{(j)} = \frac{\sum_{i=1}^{i=24} NUDIntervene_{(i,j)}}{24} \tag{6.4}$$

Ce test nous confirmera ou non l'existence d'une variation significative dans l'une des sessions.

Pour pouvoir identifier clairement la session dont la variation est significative, nous appliquons un test t (si les données suivent une distribution normale) ou un test de Mann-Whitney-Wilcoxon.

Enfin, si une variation significative est identifiée pour la troisième session nous vérifierons si celle-ci est corrélée à celle de la confiance. Pour cela, nous calculerons le cœfficient de corrélation de Spearman pour les variables $NUDInterveneMean_{(j)}$ et $TrustMean_{(j)}$, les variables $NUDInterveneMean_{(i,j)}$ et $TrustMean_{(i,j)}$, et les variables $TrustDyn_{(i,jversj+1)}$ et $NUDInterveneDyn_{(i,jversj+1)}$, avec :

$$NUDInterveneDyn_{(i,jversj+1)} = NUDIntervene_{(i,j+1)} - NUDIntervene_{(i,j)} \tag{6.5}$$

Résultats :

Le test de Shapiro-Wilk donne les résultats suivant pour la variable $NUDIntervene_{(i,j)}$:

Session	p-value
1	0,09
2	0,43
3	0,06
4	0,27

Avec un seuil de p-value à 5%, on peut considérer appliquable le test ANOVA. Nous obtenons, avec ce-dernier, une p-value de 2.236e-6. Il y a donc une différence significative dans le nombre moyen d'unité de dialogue de type intervention entre les sessions.

L'application du test t deux à deux qui nous permet de différencier les sessions deux à deux, donne le tableau de résultats suivant :

Session	1	2	3
1	-	-	-
2	3,90E-002	-	-
3	1,20E-002	3,30E-003	-
4	2,70E-002	9,05E-001	1,30E-008

Nous constatons ici aussi une différence significative (p-value $< 5\%$) de la session 3 vis-à-vis des autres. En effet l'utilisateur, lors de la panne du système, se substitue au système afin d'assurer la mission. Il est donc normal que le nombre d'unité de dialogue du type intervention varie de façon significative. Ceci confirme en partie notre hypothèse de départ. Il nous reste à vérifier que cette variation soit corrélée à celle de la confiance.

Le cœfficient de corrélation entre les variables $NUDInterveneMean_{(j)}$ et $TrustMean_{(j)}$ est de -0,94. Le cœfficient de corrélation entre les variables $NUDInterveneMean_{(i,j)}$ et $TrustMean_{(i,j)}$ est de -0,46. Enfin si nous regardons l'aspect dynamique de ces variables, soient $TrustDyn_{(i,jversj+1)}$ et $NUDInterveneDyn_{(i,jversj+1)}$, nous obtenons un cœfficient de corrélation de -0,75.

A nouveau nous avons une corrélation forte sur les variations de la confiance et du nombre d'unité de dialogue. Mais, comme pour le nombre d'unité de dialogue en général, nous prenons en compte trop d'élément à la fois. En effet, le dysfonctionnement du système a lieu sur une fonctionnalité précise.

> Lorsque la confiance diminue le nombre d'intervention augmente et inversement. Cette hypothèse est donc validée.

6.5.2 Unité de dialogue de type intervention : interception

Lors de la troisième session, le système d'interception est dysfonctionnel. Intéressons nous en premier à cet automate. Notre hypothèse de travail est que le nombre d'intervention de type "interception" est corrélé au degré de confiance globale de l'opérateur. L'évolution du nombre de commande "interception" est illustrée fig.6.8 où une forte augmentation de ce nombre est observable lors de la troisième session.

Méthode :

Nous travaillons à partir de la variable $NUDIntercept_{(i,j)}$ qui correspond au nombre d'unité de dialogue de type intervention lié à une interception. Ce nombre s'obtient en comptabilisant les unités de dialogue qui portent sur l'information "bouton ok" (validation d'une commande) avec la commande "interception" qui fait parti du champ commun (dernière commande sélectionnée).

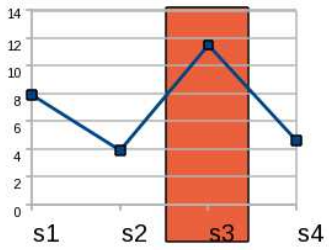

FIG. 6.8 : Evolution du nombre de commande "interception" au cours des quatre sessions. On constate une forte augmentation lors de la troisième session où le système autonome est défectueux.

Nous commencerons notre analyse avec un test Shapiro-Wilk sur cette nouvelle variable afin de déterminer si la distribution est normale ou pas. Selon le résultat nous pourrons appliquer, si la distribution est normale, un test ANOVA ainsi que des tests t deux à deux sur la variable :

$$NUDInterceptMean_{(j)} = \frac{\sum_{i=1}^{i=24} NUDIntercept_{(i,j)}}{24} \tag{6.6}$$

Sinon nous appliquerons uniquement des tests de Mann-Whitney-Wilcoxon deux à deux sur la variable $NUDInterceptMean_{(j)}$.

Enfin si nous identifions une variation significative des données lors de la troisième session, nous calculerons alors les cœfficients de corrélation de Spearman pour les variables $NUDInterceptMean_{(j)}$ et $TrustMean_{(j)}$, les variables $NUDIntercept_{(i,j)}$ et $Trust_{(i,j)}$ et les variables $TrustDyn_{(i,jversj+1)}$ et $NUDInterceptDyn_{(i,jversj+1)}$ avec :

$$NUDInterceptDyn_{(i,jversj+1)} = NUDIntercept_{(i,j+1)} - NUDIntercept_{(i,j)} \tag{6.7}$$

Résultats :

Le test de Shapiro-Wilk pour la variable $NUDIntercept_{(i,j)}$ donne les résultats suivant :

Session	p-value
1	0,09
2	0,007
3	0,52
4	0,02

Seules les sessions 1 et 3 suivent une distribution normale (p-value $> 5\%$). On utilise donc le test de Mann-Whitney-Wilcoxon qui ne fait pas d'hypothèse sur la distribution des données afin de pouvoir tester l'égalité des moyennes deux à deux. On obtient alors le tableau suivant :

Session	1	2	3
1	-	-	-
2	5,30E-003	-	-
3	2,72E-002	2,80E-005	-
4	6,60E-003	3,77E-001	6,40E-006

On constate deux regroupements : les sessions 1 et 3 d'une part et les sessions 2 et 4 d'autre part. Si nous reprenons notre protocole ce qui différencie les sessions 2, 3 et 4 est la panne du système d'interception. Nous pensons que si nous considérons uniquement ces trois sessions, nous pouvons considérer qu'il y a une variation significative liée à la panne du système d'interception concernant le nombre d'unité de dialogue pour une intervention du type interception (p-value $< 10^{-3}$).

Par contre la variation entre la première session et la seconde ne se justifie pas seulement par la prise de connaissance du but expérimental. En effet, la variation n'a une p-value que de 5% entre la session 1 et 4. C'est pourquoi nous pensons que cette variation est liée d'avantage à un phénomène d'apprentissage sur le fonctionnement de l'automate, c'est-à-dire à l'accumulation d'une plus grande expérience d'une session à la suivante.

Ces résultats semblent montrer l'existence d'une corrélation entre les variables $NUDIntercept_{(i,j)}$ et $Trust_{(i,j)}$. Nous avons une forte corrélation sur les valeurs moyennes, $NUDInterceptMean_{(j)}$ et $TrustMean_{(j)}$, -0,91 ; et faible sur les mesures, $NUDIntercept_{(i,j)}$ et $Trust_{(i,j)}$, -0,42. A nouveau, nous sommes confronté aux spécificités individuelles qui génèrent des comportements variés pour un même degré de confiance.

En ce qui concerne les variables $TrustDyn_{(i,jversj+1)}$ et $NUDInterceptDyn_{(i,jversj+1)}$, nous avons un cœfficient de corrélation de -0,69. Il existe donc un lien entre la variation de confiance et le dialogue. En effet l'opérateur qui ne fait plus confiance au système se substitue à son fonctionnement. D'un autre coté on pourrait simplement considérer que le système étant simplement en panne l'opérateur est obligé d'intervenir pour assurer la mission et qu'il n'y a pas de lien avec le changement de confiance. Or cette panne a provoqué une perte de confiance et une augmentation des interventions de l'opérateur pour les interceptions. Même si le changement stratégique au niveau du dialogue n'est pas une conséquence directe de la perte de confiance, il est tout de même une conséquence de la panne. Confiance et stratégie de dialogue subissent parallèlement des variations pour une même cause. La corrélation des deux tend à montrer que l'interventionnisme croissant de l'opérateur au sein du dialogue peut être un indicateur d'une perte de confiance.

Un deuxième commentaire est à faire sur la valeur de ce cœfficient. En ef-

fet il est légèrement plus faible que celui de la variable $TrustDyn_{(i,jversj+1)}$ avec $NUDDyn_{(i,jversj+1)}$ ou bien $NUDInterveneDyn_{(i,jversj+1)}$. Il ne faut pas oublier que la variable TrustDyn traduit une confiance globale au système. Or en analysant les unités de dialogues du type intervention par fonctionnalité, on ne regarde qu'une sous-partie du système. Cette limitation traduit donc une plus faible corrélation, mais aussi, au vu de la forte corrélation obtenue, que la confiance vis-a-vis du système globalement semble principalement lié à l'interception.

Lorsque la confiance diminue le nombre d'interception manuelle augmente (et inversement).

6.5.3 Unité de dialogue de type intervention : aller à

Nous avons vu juste avant qu'il existe un lien important entre l'évolution de la confiance, et l'intervention de l'opérateur. Regardons maintenant plus en détails comment se répartit celle-ci entre les différentes fonctionnalités du système en commençant par la commande "aller à". L'évolution du nombre de commande "aller à" est illustrée fig.6.9 où une forte augmentation de ce nombre est observable lors de la troisième session.

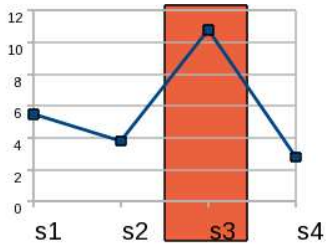

FIG. 6.9 : Evolution du nombre de commande "aller à" au cours des quatre sessions. On constate une augmentation de ce nombre lors de la troisième session où l'automate d'interception est défectueux.

Méthode :

Nous utilisons la variable $NUDGoTo_{(i,j)}$ qui est le nombre d'unité de dialogue de type intervention afin de réaliser une commande "aller à". Ce nombre s'obtient en comptabilisant les unités de dialogue qui portent sur l'information "bouton ok" (validation d'une commande) avec la commande "aller à" qui fait partie du champ commun (dernière commande sélectionnée). Afin de déterminer les outils statistiques que nous pouvons employer pour comparer les valeurs moyennes définit par la variable :

$$NUDGoToMean_{(j)} = \frac{\sum_{i=1}^{i=24} NUDGoTo_{(i,j)}}{24} \tag{6.8}$$

Nous appliquons un test de Shapiro-Wilk. Nous déterminons ainsi si la distribution de la variable $NUDGoTo_{(i,j)}$ suit une loi normale ou non. Si c'est le cas nous pourrons utiliser le test ANOVA puis le test t pour comparer les moyennes deux à deux. Sinon nous utiliserons le test de Mann-Whitney-Wilcoxon.

Enfin si nous calculons une variation significative de la variable $NUDGoToMean_{(j)}$ lors de la troisième session, nous calculerons les cœfficients de corrélation de Spearman entre les variables $TrustMean_{(j)}$ et $NUDGoToMean_{(j)}$, les variables $Trust_{(i,j)}$ et $NUDGoTo_{(i,j)}$, et les variables $TrustDyn_{(i,jversj+1)}$ et :

$$NUDGoToDyn_{(i,jversj+1)} = NUDGoTo_{(i,j+1)} - NUDGoTo_{(i,j)} \tag{6.9}$$

Résultats :

Le test de Shapiro-Wilk pour la variable $NNUDGoTo_{(i,j)}$ donne les résultats :

Session	p-value
1	0,53
2	0,59
3	0,3
4	0,009

La session 4 ne suit pas une loi normale (p-value $< 10^{-2}$). On utilisera donc le test de Mann-Whitney-Wilcoxon qui ne fait pas d'hypothèse sur la distribution des données afin de pouvoir tester l'égalité des moyennes deux à deux.

Nous obtenons alors le tableau suivant :

Session	1	2	3
1	-	-	-
2	4,30E-002	-	-
3	2,70E-004	1,30E-007	-
4	2,80E-003	3,80E-002	2,90E-009

On constate un réelle différence des données moyennes de la session 3 par rapport au trois autres (p-value $<10^{-3}$). Nous avons donc un impact de la panne du système d'interception sur des interventions de type aller à. Il est intéressant de remarquer que la commande aller à serait logiquement d'avantage lier à une remise en cause de la patrouille. Mais on peut facilement l'expliquer : la panne du système d'interception tend à bloquer les drones en vol stationnaire, par conséquence le système de patrouille n'a plus de drones pour opérer. Ainsi la panne perturbe fortement le système de patrouille. Il devient donc logique que l'opérateur ressente la nécessité d'intervenir pour redistribuer les drones immobilisés, afin d'assurer une patrouille homogène sur l'ensemble de la carte.

Le cœfficient de corrélation obtenu pour les variables $TrustMean_{(j)}$ et $NUDGoToMean_{(j)}$ est de -0,98, pour les variables $Trust_{(i,j)}$ et $NUDGoTo_{(i,j)}$ il est de -0,35, et pour les variables $TrustDyn_{(i,jversj+1)}$ et $NUDGoToDyn_{(i,jversj+1)}$ il est de -0,68. Comme nous l'avons expliqué précédemment le comportement des drones en interception à des conséquences dans leur gestion pour la patrouille. Ainsi la corrélation obtenue ici, n'est qu'un reflet des activités liées à l'interception.

> Lorsque la confiance diminue le nombre de commande "aller à" augmente car cette commande n'est pas indépendante du système d'interceptions.

6.5.4 Unité de dialogue de type intervention : ravitailler

Nous étudions, cette fois, les interventions liées à la commande "ravitailler". L'évolution du nombre de commande "ravitailler" est illustrée fig.6.10 où aucune variation particulière n'est observable lors de la troisième session.

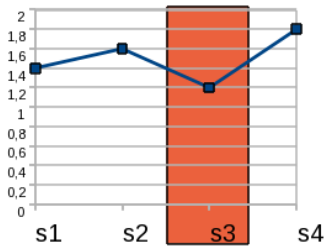

FIG. 6.10 : Evolution du nombre de commande "ravitailler" au cours des quatre sessions.

Méthode :

Nous travaillerons avec la variable $NUDRefuel_{(i,j)}$ qui est le nombre d'unité de dialogue de type intervention afin de réaliser un ravitaillement. Ce nombre s'obtient en comptabilisant les unités de dialogue qui portent sur l'information "bouton ok" (validation d'une commande) avec la commande "ravitailler" qui fait parti du champ commun (dernière commande sélectionnée).

Pour déterminer l'emploi des tests ANOVA et t ou celui de Mann-Whitney-Wilcoxon pour la variable :

$$NUDRefuelMean_{(j)} = \frac{\sum_{i=1}^{i=24} NUDRefuel_{(i,j)}}{24} \tag{6.10}$$

Nous déterminons la distribution de la variable $NUDRefuel_{(i,j)}$ à l'aide d'un test de Shapiro-Wilk. Ces tests nous permettent de déterminer une variation significative sur la valeur moyenne entre les différentes sessions.

Enfin, nous regarderons dans le cas d'une variation signification lors de la troisième session, le cœfficient de corrélation de Spearman pour les variables $TrustMean_{(j)}$ et $NUDRefuelMean_{(j)}$, les variables $Trust_{(i,j)}$ et $NUDRefuel_{(i,j)}$, et les variables $TrustDyn_{(i,jversj+1)}$ et :

$$NUDRefuelDyn_{(i,jversj+1)} = NUDRefuel_{(i,j+1)} - NUDRefuel_{(i,j)} \tag{6.11}$$

Résultats :

Le test de Shapiro-Wilk donne les résultats suivant pour la variable $NUDRefuel_{(i,j)}$:

Session	p-value
1	2,50E-005
2	1,10E-003
3	6,70E-005
4	1,30E-004

Aucune des données issues des sessions ne suit une distribution de loi normale. On utilisera donc le test de Mann-Whitney-Wilcoxon. On obtient alors le tableau suivant :

Session	1	2	3
1	-	-	-
2	0,59	-	-
3	0,8	0,3	-
4	0,46	0,72	0,13

Les moyennes de toutes les sessions peuvent être considérer comme égales. Il n'y a donc pas de variation significative dans les interventions de type ravitailler. En effet, contrairement à la fonction de patrouille, la fonction de ravitaillement n'est pas perturbée par la panne lors de la session 3 parce qu'elle est prioritaire à l'exécution : si un drone en interception doit aller ravitailler, il le fera. Le comportement apparent des drones associé à cette fonction est donc le même dans toutes les sessions. Il paraît donc logique que le comportement de l'opérateur ne change pas de manière significative d'une session à l'autre.

Lorsque la confiance diminue le nombre de commande "ravitailler" n'évolue pas car cette commande est indépendante du système d'interceptions.

6.5.5 Unité de dialogue de type intervention : patrouiller

Étudions maintenant le dernier type d'intervention, la commande "patrouiller". L'évolution du nombre de commande "patrouiller" est illustrée fig.6.11 où aucune variation particulière n'est observable lors de la troisième session.

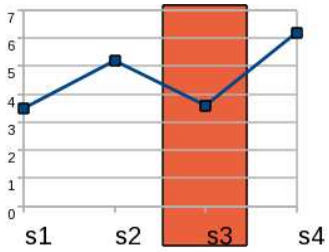

FIG. 6.11 : Evolution du nombre de commande "patrouiller" au cours des quatre sessions.

Méthode :

Nous définissons la variable $NUDPatrol_{(i,j)}$ comme le nombre d'unité de dialogue de type intervention afin de réaliser une remise en patrouille. Ce nombre s'obtient en comptabilisant les unités de dialogue qui portent sur l'information "bouton ok" (validation d'une commande) avec la commande "patrouiller" qui fait parti du champ commun (dernière commande sélectionnée).

Nous utilisons ensuite un test de Shapiro-Wilk pour vérifier que la distribution de la variable suit une loi normale. Nous pourrons ainsi utiliser le test ANOVA et le test t pour comparer, entre les différentes sessions, la variable :

$$NUDPatrolMean_{(j)} = \frac{\sum_{i=1}^{i=24} NUDPatrol_{(i,j)}}{24} \tag{6.12}$$

Si $NUDPatrol_{(i,j)}$ ne suit pas une distribution normale, nous utiliserons alors le test de Mann-Whitney-Wilcoxon pour comparer deux à deux les moyennes de chaque session.

Enfin l'observation d'une variation significative de la variable $NUDPatrolMean_{(j)}$, nous amène à vérifier les corrélations entre les variables $TrustMean_{(j)}$ et $NUDPatrolMean_{(j)}$, les variables $Trust_{(i,j)}$ et $NUDPatrol_{(i,j)}$, et les variables $TrustDyn_{(i,jversj+1)}$ et :

$$NUDPatrolDyn_{(i,jversj+1)} = NUDPatrol_{(i,j+1)} - NUDPatrol_{(i,j)} \tag{6.13}$$

Le cœfficient utilisé est celui de Spearman.

Résultats :

Nous obtenons les résultats suivant pour le test Shapiro-Wilk avec la variable $NUDPatrol_{(i,j)}$:

Session	p-value
1	8,50E-008
2	1,90E-004
3	5,10E-005
4	1,20E-004

Aucune des données issues des sessions ne suit une distribution de loi normale. On utilisera donc le test de Mann-Whitney-Wilcoxon qui ne fait pas d'hypothèse sur la distribution des données afin de pouvoir tester l'égalité des moyennes deux à deux. Nous avons alors le tableau suivant :

Session	1	2	3
1	-	-	-
2	3,73E-001	-	-
3	9,74E-001	1,87E-001	-
4	1,42E-001	4,81E-001	1,30E-002

L'intervention de type patrouiller à pour but d'annuler des ordres émis par l'opérateur ou le système (interception, ravitaillement, aller à). Il est intéressant de noter que la panne du système d'interception qui tend à immobiliser les drones inutilement n'a pas eu de conséquence sur l'usage de cette commande. Afin d'assurer la mission de surveillance il faut donc remettre les drones en mouvement. Pour cela la solution immédiate est d'utiliser la commande "patrouiller". Or nous constatons que lors du troisième scénario, les dialogues concernant cette commande ne varient pas significativement. Les opérateurs privilégient alors deux autres solutions :

- La correction d'une interception manuellement. C'est-à-dire que l'interception automatique du système est annulée par une commande d'interception de l'opérateur. Il y a une action de correction. Ce qui signifie, contrairement à ce que nous avons pu dire précédemment, que l'opérateur ne se substitue pas forcément au système, il peut tout simplement se contenter de le corriger.
- L'annulation de la commande avec une compensation du système de patrouille. En effet ce dernier, ayant de moins en moins de drones, ne peut plus assurer une surveillance complète de la carte. L'opérateur lorsqu'il remet les drones en mouvement, utilise la commande "aller à" pour les redistribuer sur l'ensemble de la carte. Ce qui est parfaitement en accord avec le résultat vu précédemment.

Lorsque la confiance diminue le nombre de commande "patrouiller" n'évolue pas car cette commande n'est pas liée au système d'interception.

Nous avons vérifié ici que notre modèle de dialogue permet d'observer, avec une quantification des dialogues de la couche intervention, une catégorie de contrôle exercé par l'opérateur. En effet, selon Castelfranchi, la notion de contrôle implique d'une part la vérification de la réalisation d'une tâche, et d'autre part les actions qui permettent d'assurer cette réalisation. Ainsi, la corrélation obtenue entre les variations de la confiance et les variations du nombre d'intervention de l'opérateur montre que ce second aspect du contrôle est observable au travers de notre modèle. De plus, lorsque l'on tient compte du contenu sémantique (information d'interception, de ravitaillement ou de patrouille) on constate que la définition d'objet de confiance, présentée en introduction du second chapitre, prend tout son sens. En effet, nous avons constaté que l'évolution du comportement de l'opérateur était réellement ciblé aux fonctionnalités mises en défaut.

En ce qui concerne le premier aspect de la notion de contrôle, à savoir le suivi d'une tâche, il faut nous intéresser aux rouages internes de notre modèle de dialogue.

6.6 Etude du lien entre confiance et transition intra-dialogue

Pour le moment nous avons abordé une mesure au sein de notre modèle qui reflète le contrôle comme ajustement d'un processus. Le second aspect du contrôle est le suivi du processus. Pour cela, nous avons mis en avant dans notre modèle un certain nombre de mécanismes de contrôle. En effet nous considérons la relation entre confiance et contrôle comme une relation d'opposition, on doit donc pouvoir au travers de notre modèle observer une corrélation entre les mesures de confiance et une quantification des mécanismes de contrôle mis en jeu au cours du dialogue. C'est pourquoi nous proposons maintenant d'étudier ce lien avec comme élément de contrôle l'acquittement, les demandes d'explications et les demandes d'extension de l'information.

6.6.1 Demande d'explication

Commençons notre étude avec la transition "reqExplain(op)" qui, dans les unités de dialogue de type monitoring, est activé lorsque l'opérateur formule une demande d'explication. Nous supposons qu'un lien existe entre la confiance de l'opérateur et la quantité d'explication qu'il demande au système. L'évolution du nombre de demande d'explication est illustrée fig.6.12 où aucune variation particulière n'est observable lors de la troisième session.

Méthode :

Pour notre étude nous définissons la variable $NReqExpl_{(i,j)}$ qui est le nombre de transition "reqExplain" au cours d'un dialogue de type monitoring pour le sujet i dans la session j. Cela correspond au nombre de fois que l'utilisateur j a cliqué sur un bouton '?' au sein des pop-up d'informations.

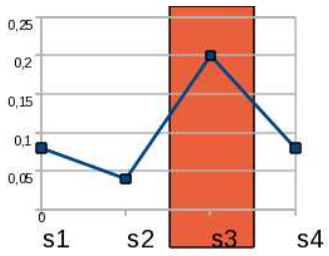

FIG. 6.12 : Evolution du nombre de demande d'explication au cours des quatre sessions.

Nous emploierons la même démarche que précédemment. Nous commençons par vérifier la distribution de notre variable afin de déterminer les tests que comparaison que nous emploierons ensuite. Ainsi un test de Shapiro-Wilk est utilisé en premier puis selon la réponse un test ANOVA compléter de tests t (distribution normale de notre variable) ou des tests Mann-Whitney-Wilcoxon. Ces derniers nous permettrons d'établir l'existence ou non de variation significative d'une session expérimentale à l'autre de la variable :

$$NReqExplMean_{(j)} = \frac{\sum_{i=1}^{i=24} NReqExpl_{(i,j)}}{24} \tag{6.14}$$

Enfin dans le cadre d'une variation lors de la troisième session, nous calculerons les cœfficients de corrélation de Spearman des variables $TrustMean_{(j)}$ et $NReqExplMean_{(j)}$, les variables $Trust_{(i,j)}$ et $NReqExpl_{(i,j)}$, et les variables $TrustDyn_{(i,jversj+1)}$ et :

$$NReqExplDyn_{(i,jversj+1)} = NReqExpl_{(i,j+1)} - NReqExpl_{(i,j)} \tag{6.15}$$

Résultats :

Nous obtenons, pour le test de Shapiro-Wilk les résultats suivant :

Session	p-value
1	1,40E-009
2	2,20E-010
3	1,60E-009
4	2,20E-010

Aucune des données issues des sessions ne suit une distribution de loi normale. On utilisera donc le test de Mann-Whitney-Wilcoxon qui ne fait pas d'hypothèse sur la distribution des données afin de pouvoir tester l'égalité des moyennes deux à deux. On obtient le tableau suivant :

Session	1	2	3
1	-	-	-
2	0,33	-	-
3	0,33	0,16	-
4	1	0,66	0,48

Nous n'avons donc aucune variation significative sur le nombre de demande d'explication entre les sessions. Il faut noter que cette fonction est très peu utilisée. Les utilisateurs ont été formés à l'utilisation du système avec un scénario d'apprentissage de 10 minutes, toutes les fonctionnalités n'ont pas forcément été mémorisées par le sujet d'autant plus que certaines fonctionnalités telles que la demande d'information peuvent paraître sans intérêt au vu de la complexité de la tâche. Ainsi les sujets n'ont pas ressenti le besoin d'approfondir leur connaissance sur le comportement du système pour prendre la décision d'agir. Le simple fait de voir le système accomplir sa tâche ou non correctement est, selon nous, un critère suffisant du point de vue des opérateurs pour prendre des décisions.

L'hypothèse de relation d'opposition entre le nombre de demandes d'explication et le degré de confiance de l'opérateur n'est pas vérifiée.

6.6.2 Demande d'information étendue

Intéressons nous maintenant au deuxième mécanisme de contrôle mis en exergue dans notre modèle de dialogue : les demandes d'information étendue. L'évolution du nombre de demande d'information étendue est illustrée fig.6.13 où aucune variation particulière n'est observable lors de la troisième session.

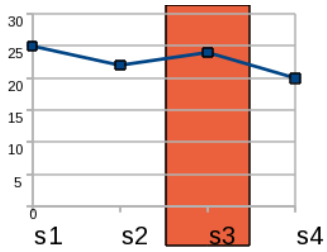

FIG. 6.13 : Evolution du nombre de demande d'information étendue au cours des quatre sessions.

Méthode :

Notre étude est réalisé à partir de la variable $NReqExt_{(i,j)}$ qui est le nombre de transition "reqExtend" au cours d'un dialogue de type monitoring pour le sujet i

dans la session j. Cela correspond au nombre de fois où l'utilisateur j a ouvert des pop-up d'informations et où il a cliqué sur un bouton '+' au sein celles-ci.

Nous étudions en premier sa distribution à l'aide d'un test de Shapiro-Wilk. Si celle-ci suit une loi normale nous pourrons ensuite appliquer un test ANOVA et des tests t sur la variable :

$$NReqExtMean_{(j)} = \frac{\sum_{i=1}^{i=24} NReqExt_{(i,j)}}{24} \tag{6.16}$$

Sinon nous utiliserons des tests Mann-Whitney-Wilcoxon. Ces tests nous permettrons de comparer la quantité moyenne de demande d'information étendue entre chaque session. Ainsi nous pourrons vérifier ou non l'existence d'une variation lors de la troisième session en corrélation avec la confiance. C'est pourquoi, dans le cas d'une variation significative, nous calculerons les cœfficients de corrélation de Spearman pour les variables $TrustMean_{(j)}$ et $NReqExtMean_{(j)}$, les variables $Trust_{(i,j)}$ et $NReqExt_{(i,j)}$, et les variables $TrustDyn_{(i,jversj+1)}$ et :

$$NReqExtDyn_{(i,jversj+1)} = NReqExt_{(i,j+1)} - NReqExt_{(i,j)} \tag{6.17}$$

Résultats :

Nous obtenons, pour le test de Shapiro-Wilk les résultats suivant :

Session	p-value
1	4,90E-002
2	6,70E-003
3	1,20E-002
4	1,80E-001

Aucune des données issues des sessions ne suit une distribution de loi normale. On utilisera donc le test de Mann-Whitney-Wilcoxon qui ne fait pas d'hypothèse sur la distribution des données afin de pouvoir tester l'égalité des moyennes deux à deux. Nous avons alors le tableau de résultats suivant :

Session	1	2	3
1	-	-	-
2	0,55	-	-
3	0,79	0,72	-
4	0,18	0,54	0,31

De la même façon que précédemment, les fonctionnalités rattachées à ce type de transition n'ont été que peu utilisées. De plus, leur utilisation a été principalement la consultation du niveau du carburant des drones. En effet celui-ci n'était pas

accessible directement sur l'interface. Enfin l'automate qui gère le ravitaillement était fonctionnel lors de toutes les sessions, il est donc normal que le comportement des sujets vis-à-vis du fuel ne varie pas.

L'hypothèse de relation d'opposition entre le nombre de demandes d'informations étendues et le degré de confiance de l'opérateur n'est pas vérifiée.

6.6.3 Acquittement

Regardons le troisième mécanisme de contrôle que nous avons décrit : l'acquittement. Ce dernier selon sa fréquence traduira une attention de l'opérateur plus ou moins conséquente sur l'information. Nous supposons que plus l'opérateur prend connaissance de l'information moins il a confiance. Nous étudierons donc ici le nombre d'acquittements effectués au cours des sessions en relation avec la confiance de l'opérateur. L'évolution du nombre d'acquittements est illustrée fig.6.14 où on observe une augmentation de ces derniers.

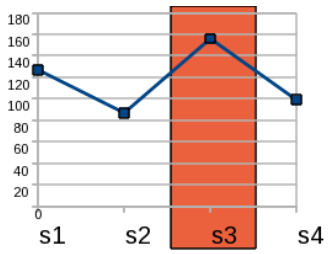

FIG. 6.14 : Evolution du nombre d'acquittements (réalisés par focus attentionel) au cours des quatre sessions. On constate une augmentation de ces derniers lors de la troisième session.

Méthode :

Nous définissons la variable $NAck_{(i,j)}$ comme le nombre de transition "ack" au cours d'un dialogue de type monitoring pour le sujet i dans la session j. Cela correspond au nombre de fois où l'utilisateur j a focalisé son attention sur un élément information (un focus attentionnel long est comptabilisé plusieurs fois selon la fréquence d'échantillonnage de la mesure du focus).

Nous étudions en premier la distribution de cette variable à l'aide d'un test de Shapiro pour déterminer si elle suit une loi normale ou non. Dans le cas d'une distribution normale nous pourrons utiliser par la suite les test ANOVA et t pour étudier les variations de la variable suivante entre les sessions :

$$NAckMean_{(j)} = \frac{\sum_{i=1}^{i=24} NAck_{(i,j)}}{24} \tag{6.18}$$

Dans le cas d'une distribution autre de la variable $NAck_{(i,j)}$, nous utiliserons alors le test de Mann-Whitney-Wilcoxon pour les comparaisons de moyenne.

Dans tout les cas si une variation significative est identifiée lors de la troisième session nous calculerons alors le cœfficient de corrélation de Spearman entre les variables $TrustMean_{(j)}$ et $NAckMean_{(j)}$, les variables $Trust_{(i,j)}$ et $NAck_{(i,j)}$, et les variables $TrustDyn_{(i,jversj+1)}$ et :

$$NAckDyn_{(i,jversj+1)} = NAck_{(i,j+1)} - NAck_{(i,j)} \tag{6.19}$$

Résultats :

Nous obtenons, pour le test de Shapiro-Wilk les résultats suivant :

Session	p-value
1	6,20E-002
2	0,4
3	3,80E-002
4	7,80E-002

Aucune des données issues des sessions ne suit une distribution de loi normale. On utilisera donc le test de Mann-Whitney-Wilcoxon qui ne fait pas d'hypothèse sur la distribution des données afin de pouvoir tester l'égalité des moyennes deux à deux. Nous avons alors les résultats suivant :

Session	1	2	3
1	-	-	-
2	3,70E-003	-	-
3	2,83E-002	8,00E-008	-
4	2,47E-002	1,40E-001	1,10E-007

Nous avons une variation significative du nombre d'acquittement au cours de la troisième session. L'opérateur surveille d'avantage les éléments informatifs du système. La panne du système d'interception accapare d'avantage l'attention du sujet. Si nous regardons les cœfficients de corrélation de Spearman entre les variables :

- $NAckMean_{(j)}$ et $TrustMean_{(j)}$, on obtient -1 ;
- $NAck_{(i,j)}$ et $Trust_{(i,j)}$, on obtient -0,41 ;
- $NAckDyn_{(i,jversj+1)} = NAck_{(i,j+1)} - NAck_{(i,j)}$, et $TrustDyn_{(i,jversj+1)}$, on obtient -0,73.

Nous avons donc une corrélation forte entre la variation de la confiance et la variation du nombre d'acquittement de l'opérateur. Comme nous l'avons précisé précédemment, les acquittements dans les dialogues de type monitoring sont liés au focus attentionnel de l'opérateur. On conclut donc que l'opérateur a une attention plus importante lorsqu'il ne fait pas confiance au système.

L'hypothèse de relation d'opposition entre le nombre d'acquittement par focus attentionel et le degré de confiance de l'opérateur est vérifiée.

Nous avons voulu vérifier, à travers cette étude, si les mécanismes de contrôle mis en œuvre au sein de notre modèle permettait d'évaluer la confiance d'un opérateur vis-à-vis de son système. Pour cela nous partont du principe que confiance et contrôle sont liés par une relation d'opposition et qu'en quantifiant les contrôles de l'opérateur sur le système, nous avons une indication du degré de confiance. Or bien que l'on est observé une corrélation entre les variations du focus attentionnel de l'opérateur (qui déclenche une transition de d'acquittement) et de la confiance, les mécanismes d'extension de l'information et d'explication n'ont pas du tout été sollicités par les sujets. Nous ne validons ici que partiellement notre modèle. En effet nous pensons que la simplicité de la tâche ne demandait aucun effort de compréhension de la part de l'opérateur. La simple observation du système (qui induit par ailleurs l'augmentation des acquittements lors de la troisième session) pour voir si la tâche est accomplie ou non suffit à l'opérateur pour décider d'agir ou non.

6.7 Etude de la cohérence entre les deux évaluations de confiance

Notre approche de l'évaluation de la confiance par le dialogue est motivée par une nécessité d'anticipation des situations de mauvais usage d'un automate en raison d'un mauvais niveau de confiance. Ainsi, les résultats précédemment obtenus à l'échelle d'une session, sont d'autant plus intéressant si nous les retrouvons sur une échelle de temps plus courte. Pour pouvoir faire cette étude nous allons utiliser les évaluations de confiance obtenues lors du visionnage des sessions. Cette évaluation consiste donc a une auto-évaluation du sujet au cours d'un re-jeu et elle est liée aux évènements du scénario.

Mais avant d'exploiter les données sur la confiance issue du re-jeu, nous souhaitons d'abord vérifier leur cohérence avec le questionnaire. Pour cela nous allons rechercher l'existence d'une corrélation entre ces méthodes. Puis nous vérifierons la cohérence des mesures obtenues par auto-évaluation sur les sessions successives, à savoir si le degré de confiance entre la fin d'une session et le début de la session suivante est cohérente.

6.7.1 Etude de la cohérence mutuelle des mesures

L'objectif de cette analyse est de montrer l'existence d'une corrélation forte entre les deux méthodes utilisées pour l'évaluation de la confiance. En effet, comme le montrent les courbes fig.6.15 et fig.6.16, ces deux modes d'évaluations montrent des résultats visuellement comparables : on observe dans les deux cas une forte perte de confiance des opérateurs.

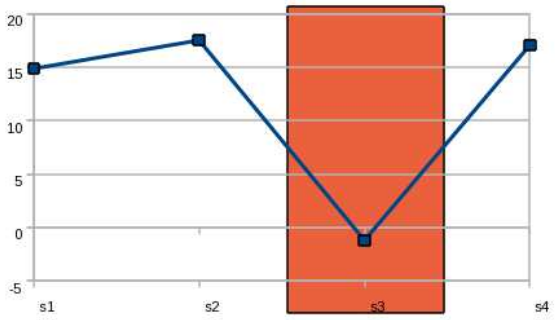

FIG. 6.15 : Evolution de la confiance au cours des quatre sessions. Le degré de confiance mesuré est obtenu à l'aide du questionnaire de Jian. On constate la perte de confiance lors de la troisième session.

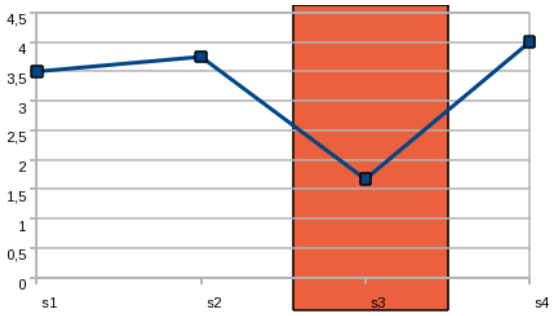

FIG. 6.16 : Evolution de la confiance au cours des quatre sessions. Le degré de confiance mesuré est obtenu à l'aide par une auto-évaluation du sujet sur une échelle de 1 à 5. On constate la perte de confiance lors de la troisième session.

Méthode :

Pour l'évaluation de la confiance obtenue par auto-évaluation, nous définissons la variable :

– $TrustRMean_{(i,j)} =$

$$\frac{\sum TrustR_{(i,j,n)} * (TrustRTime_{(i,j,n+1)} - TrustRTime_{(i,j,n)})}{10} \tag{6.20}$$

avec $TrustR_{(i,j,n)}$ l'auto-évaluation de confiance du sujet i à la session j pour le n-ième évènement, $TrustRTime_{(i,j,n)}$ le temps écoulé depuis le début de la session j et le tout diviser par la durée d'une session, soit 10 minutes.

Nous calculons ensuite les cœfficients de Spearman entre les variables $Trust_{(i,j)}$ (évaluation issue du questionnaire) et $TrustRMean_{(i,j)}$ (confiance moyenne issue du re-jeu), puis entre les variables $TrustDyn_{(i,jversj+1)}$ et :

$$TrustRMeanDyn_{(i,jversj+1)} = TrustRMean_{(i,j+1)} - TrustRMean_{(i,j)} \tag{6.21}$$

Résultats :

Le cœfficient de corrélation entre les variables $Trust_{(i,j)}$ et $TrustRMean_{(i,j)}$ est de 0,62. Nous avons donc une corrélation forte entre les valeurs tendant à montrer que les données obtenues au cours du re-jeu sont cohérentes avec celles du questionnaire. Ceci est d'autant plus vrai lorsque nous nous intéressons cette fois aux variations entre les sessions. En effet le cœfficient de corrélation des variables $TrustDyn_{(i,jversj+1)}$ et $TrustRMeanDyn_{(i,jversj+1)}$ est alors de 0,8. On peut donc considérer que l'utilisation des évaluations de re-jeu, de par leur cohérence avec les données issues du questionnaire, est possible pour analyser l'évolution de la confiance au sein d'une session.

Les deux méthodes d'évaluation subjective de la confiance sont cohérentes entre elles.

6.7.2 Etude de la cohérence de l'auto-évaluation des sessions successives

Nous savons que les résultats obtenus à l'aide des deux méthodes d'évaluation sont comparables au niveau des moyennes et de leur variation. Afin de renforcer la validité de l'auto-évaluation, nous allons à présent vérifier la cohérence des mesures d'auto-évaluation entre les débuts et fins de session. L'idée est de regarder si d'une session à l'autre l'auto-évaluation finale d'une session j est cohérente avec la première auto-évaluation de la session j+1. Nous vérifions aussi que la perte ou le gain de confiance d'une session j à la suivante à bien lieu au sein de la session j.

Méthode :

Pour commencer, nous définissons les variables :
- $TrustRBegin_{(i,j)}$ est l'auto-évaluation du sujet i de sa confiance a priori de la session j.
- $TrustREnd_{(i,j)}$ est la dernière auto-évaluation du sujet i à la session j.

Afin de comparer les fins et débuts de session, nous regardons le cœfficient de corrélation entre les variables $TrustREnd_{(i,j)}$ et $TrustRBegin_{(i,j+1)}$:

Résultats :

L'idée est de regarder si d'une session à l'autre l'auto-évaluation finale d'une session j est cohérente avec la première auto-évaluation de la session j+1.

Session	Cœfficient de corrélation de Spearman
1 vers 2	0,56
2 vers 3	0,77
3 vers 4	0,54

Nous avons des corrélations fortes entre chacune des sessions, mais peu convaincante de la session 1 vers 2 et 3 vers 4. Avant d'analyser ce point faisons un test de comparaison de moyenne entre les variables :

$$TrustREndMean_{(j)} = \frac{\sum_{i=1}^{i=24} TrustREnd_{(i,j)}}{24} \tag{6.22}$$

$$TrustRBeginMean_{(j)} = \frac{\sum_{i=1}^{i=24} TrustRBegin_{(i,j)}}{24} \tag{6.23}$$

Nous comparons les couples de moyenne suivants :

Session	$\boldsymbol{TrustREndMean_{(j)}}$	$\boldsymbol{TrustRBeginMean_{(j+1)}}$
1 vers 2	3,5	3,9
2 vers 3	3,75	4,2
3 vers 4	1,7	2,6

Nous appliquons un test de Mann-Whitney-Wilcoxon car les distributions de ces deux variables (pour chaque valeur de j) ne sont pas normales.

Session	p-value
1 vers 2	0,15
2 vers 3	0,19
3 vers 4	0,008

TAB. 6.1 :

Les moyennes n'ont aucune variation significative entre les sessions 1 et 2, et 2 et 3. Ceci est tout à fait logique, dans le sens où l'utilisateur se construit un historique relationnel positif avec le système. Il y a donc une continuité du degré de confiance d'une session à l'autre.

Cette continuité est brisée à la fin de la troisième session. En effet, la panne de l'automate introduit un retour d'expérience négatif par rapport aux sessions précédentes. Le degré de confiance des sujets est très bas à la fin de la troisième session. On peut donc penser que le démarrage d'une nouvelle session, leur fait réévaluer leur confiance en prenant du recul sur la troisième session. En effet, durant deux sessions le système fonctionnait correctement. Ainsi les sujets commencent la session 4 avec une faible confiance mais pas aussi basse qu'elle l'a été à la fin de la troisième session.

Nous avons vu précédemment une variation non négligeable de la variable $Trust_{(i,j)}$ entre les sessions 2 et 3, et 3 et 4. Si nous regardons plus attentivement nous nous rendons compte que cette évolution à lieu durant la session 3 (pour la différence 2 et 3), et durant la session 4 (pour la différence 3 et 4).

Nous avons vu juste avant (3.1.2) que l'évaluation de la confiance entre la fin de la session 2 et le début de la session 3 est fortement corrélée. Or si nous comparons les moyennes entre le début et la fin de la session 2, elles sont comparables (pas de variations significatives avec un test de Mann-Whitney-Wilcoxon). Cela implique donc que l'évolution a eu lieu uniquement durant la session 3 (variations significative avec un test de Mann-Whitney-Wilcoxon).

Session	p-value
2	0,59
3	2.449e-08

La comparaison des moyennes entre le début de la session 3 et la fin de la session montre une variation significative. Leur valeur montre que cette variation est à la baisse (on passe de 4,2 à 1,7, fig.6.17). C'est à dire que la perte de confiance se fait durant la troisième session. Si nous prenons quelques exemples nous pouvons voir que cette perte de confiance est progressive et plus ou moins rapide selon les profils individuels. L'évaluation basée sur une échelle de 1 à 5 par valeur entière, le sujet 1, par exemple, passe d'une évaluation initiale de 4 à une évaluation de 1 en 3 min 46 s. Il est à noter que trois sujets font exceptions : leur degré de confiance ne varie pas entre le début et la fin de la session. Aucun sujet ne voit sa confiance croître.

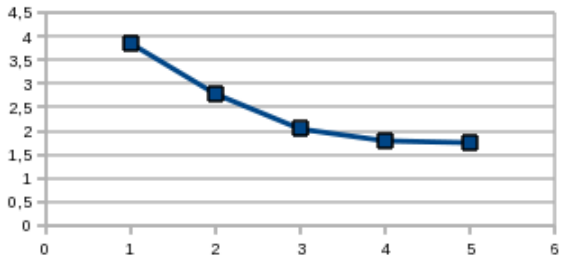

FIG. 6.17 : Evolution de la confiance au cours de la troisième session. Celle-ci a été divisé en cinq segments de 2 minutes. On constate la diminution progressive de la confiance lors de cette session.

Regardons maintenant l'évolution de la confiance lors de la quatrième session. Cette fois la confiance est toujours à la hausse entre le début et la fin de la session (fig.6.18). La variation est significative lorsqu'on applique un test de Mann-Whitney-Wilcoxon entre la moyenne initiale (2,6) et la moyenne finale (4).

Session	p-value
4	0,59
3	1.184e-04

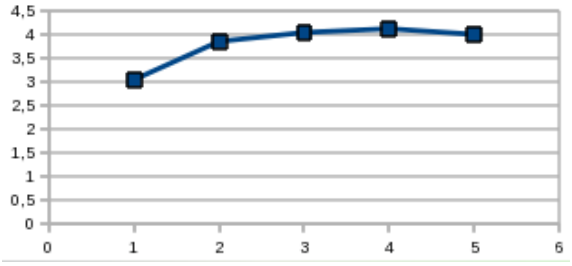

FIG. 6.18 : Evolution de la confiance au cours de la quatrième session. Celle-ci a été divisé en cinq segments de 2 minutes. On constate le rétablissement progressif de la confiance lors de cette session.

La mesure subjective de la confiance au cours du re-jeu est cohérente d'une session à l'autre.

6.8 ETUDE DU LIEN ENTRE CONFIANCE ET NOMBRE D'UNITÉ DE DIALOGUE : ANALYSE INTRA-SESSION

Notre expérimentation est très dense en activité et en échange d'informations. Nous pouvons, grâce au résultat précédent, réaliser nos statistiques dans un intervalle de temps inférieur à la durée d'une session (soit inférieur à 10 minutes). Notre modèle dépend de la quantité de dialogue entre l'opérateur et le système. De plus, cette réduction de la fenêtre temporelle est nécessaire dans ce type de contexte. En effet, la grande quantité d'informations qui circulent amène l'opérateur à réévaluer sa confiance plus fréquemment. Ainsi si nous voulons réellement étudier la co-évolution de la confiance et des mécanismes de contrôle, il nous faut réduire l'historique de dialogue. Le période de temps à choisir dépend du contexte. Pour cette étude nous prendrons les deux dernières minutes de dialogue pour l'historique. Reprenons maintenant notre étude pour les parties de notre modèle que nous avons validées précédemment (intervention, acquittement).

Jusqu'à présent nous supposons que confiance et contrôle sont liés par opposition et nous cherchons des corrélations entre ces deux notions aux travers de notre modèle. En réduisant notre échelle de temps, nous voulons montrer que l'évolution de la confiance n'est pas immédiate. En effet, lorsqu'un opérateur contrôle son système, il exprime un doute vis-à-vis de ce dernier. Ce contrôle a pour conséquence, selon les conclusions de l'opérateur, de diminuer sa confiance ou non. Ce processus implique alors un délai entre les premiers contrôles de l'opérateur et sa perte effective de confiance. Ainsi l'observation de la mise en œuvre de mécanisme de contrôle par l'opérateur au sein de notre modèle n'est pas synonyme de méfiance tant que cela reste ponctuel. C'est le maintien d'une stratégie de contrôle observée au travers de notre modèle qui impliquera une diminution de la confiance de l'opérateur. C'est ce que nous proposons de voir pour les mécanismes de contrôle déjà mis en corrélation

avec la confiance (au niveau inter-session).

6.8.1 Confiance

Dans un premier temps, vérifions que l'évolution de la confiance d'un segment à l'autre (d'une période à la suivante) est en accord avec nos conclusions de la section précédente.

Méthode :

Nous allons vérifier dans un premier temps la distribution des données pour voir si elles suivent une loi normale. Pour cela nous utilisons le test de Shapiro-Wilk. Nous appliquons donc ce test à la variable $TrustRMean_{(j,t)}$ qui représente la confiance moyenne, au cours du segment t, mesurée par la méthode du re-jeu.

Ensuite nous comparerons à l'aide d'un test de Mann-Whitney-Wilcoxon, ou à l'aide d'un test t (selon la distribution) pour vérifier si les variations d'un segment à l'autre sont significatifs.

Résultats :

Nous obtenons les résultats suivant pour le test Shapiro-Wilk avec la variable $TrustRMean_{(j,t)}$:

Segment	p-value
1	8,9E-02
2	3,1E-01
3	5,9E-04
4	7,6E-05
5	1,2E-05

Les données des segments 3 à 5 ne suivent pas une loi normale. Nous utiliserons donc le test de Mann-WHitney-Wilcoxon pour comparer les sessions successives 2 à 5, et le test t pour comparer les deux premières sessions. On obtient alors le tableau suivant :

Segments	p-value
1 vers 2	1,1E-03
2 vers 3	3,7E-02
3 vers 4	5,3E-01
4 vers 5	5,9E-01

On constate alors que les trois premiers segments temporels voient décroître significativement le degré de confiance de l'opérateur, en accords avec les résultats

présentés à la section 6.7.2. Au vu des résultats présentés précédemment regardons si nos indicateurs de confiance suivent une même progression (les variations de confiance et de contrôle étant corrélées). Ce résultat ne serait pas logique puisque la perte de confiance de l'opérateur est lié à une réévaluation de sa confiance suite aux contrôles effectués par l'opérateur. Nous nous attendons plutôt à un délai entre la mise en œuvre d'une stratégie de contrôle de l'opérateur et l'évolution de sa confiance.

La perte de confiance est progressive au cours de la troisième session.

6.8.2 Intervention

Nous avons montré que notre modèle nous permettait d'observer les mécanismes de contrôle de type intervention. L'évolution du nombre d'intervention était corrélé à l'évolution de la confiance d'une session à une autre. Nous allons maintenant regarder si la corrélation est toujours vraie lorsque l'on analyse leurs évolutions au sein de la troisième session.

Méthode :

Pour cette étude nous définissons une nouvelle variable : $NbUDInterveneEvol_{(i,j,t)}$. Elle représente le nombre d'unité de dialogue de type intervention (initiée par l'opérateur) au cours du segment temporel t. Ce nombre s'obtient en comptabilisant les unités de dialogue qui portent sur l'information "bouton ok" (validation d'une commande) pendant un segment temporel.

Dans un premier temps nous allons vérifier l'existence d'une variation significative de cette variable entre deux segments temporels consécutifs. Pour cela nous commencerons par vérifier la distribution de $NbUDInterveneEvol_{(i,j,t)}$ à l'aide d'un test de Shapiro-Wilk. En effet, cela nous permettra de savoir si nous pouvons appliquer un test ANOVA sur la variable :

$$NbUDInterveneEvolMean_{(j,t)} = \frac{\sum_{i=1}^{i=24} NbUDInterveneEvol_{(i,j,t)}}{24} \tag{6.24}$$

Ce test nous confirmera ou non l'existence d'une variation significative dans l'un des segments temporels.

Pour pouvoir identifier clairement la session dont la variation est significative, nous appliquons un test t (si les données suivent une distribution normale) ou un test de Mann-Whitney-Wilcoxon.

Enfin, si une variation significative est identifiée nous vérifions si celle-ci est corrélée à celle de la confiance. Pour cela, nous calculerons le cœfficient de corrélation de Spearman pour les variables $NbUDInterveneMeanEvol_{(j,t)}$ et $TrustRMean_{(j,t)}$.

Résultats :

Nous obtenons les résultats suivant pour le test Shapiro-Wilk avec la variable $NbUDPatrol_{(i,j)}$:

Segment	p-value
1	5,8E-03
2	1,2E-02
3	3,9E-03
4	9,0E-03
5	1,4E-02

Aucune des données issues des sessions ne suit une distribution de loi normale. On utilisera donc le test de Mann-Whitney-Wilcoxon qui ne fait pas d'hypothèse sur la distribution des données afin de pouvoir tester l'égalité des moyennes deux à deux. Nous avons alors le tableau suivant :

Session	1	2	3	4
1	-	-	-	-
2	0,45	-	-	-
3	0,96	0,38	-	-
4	0,74	0,21	0,69	-
5	0,69	0,69	0,73	0,38

Nous ne constatons aucune différence significative du nombre d'interventions au cours de la session. Il est normal à partir du moment où le système ne fonctionne plus correctement qu'il devient nécessaire pour l'opérateur d'agir. Or nous pensons que c'est le maintien d'une telle situation qui va petit à petit rendre l'opérateur méfiant envers les actions du système. Ceci implique alors que la confiance évolue avec un temps de réponse à la mise en œuvre d'une stratégie de contrôle par l'opérateur. Ainsi plus la situation se maintient, plus la confiance de l'opérateur diminue.

> Sur une échelle de temps courte, l'évolution du degré de confiance et du nombre d'intervention n'est plus en opposition. Le degré de confiance évolue avec un temps de réponse au changement de stratégie de l'opérateur.

6.8.3 Acquittement

Après l'intervention, intéressons nous à la seconde partie validée de notre modèle, à savoir l'évaluation de la confiance à travers la quantification des acquittements de l'opérateur lors des dialogues de type suivi. Nous pensons établir un résultat identique au précédent.

Méthode :

Pour cette étude nous définissons une nouvelle variable : $NbAckEvol_{(i,j,t)}$. Elle représente le nombre d'acquittement réalisé par l'opérateur au cours du segment temporel t. Ce nombre s'obtient en comptabilisant les transitions "ack(op)" pendant un segment temporel.

Dans un premier temps nous allons vérifier l'existence d'une variation significative de cette variable d'un segment temporel au suivant. Pour cela nous commencerons par vérifier la distribution de $NbAckEvol_{(i,j,t)}$ à l'aide d'un test de Shapiro-Wilk. En effet, cela nous permettra de savoir si nous pouvons appliquer un test ANOVA sur la variable :

$$NbAckEvolMean_{(j,t)} = \frac{\sum_{i=1}^{i=24} NbAckEvol_{(i,j,t)}}{24} \tag{6.25}$$

Ce test nous confirmera ou non l'existence d'une variation significative dans l'un des segments temporels.

Pour pouvoir identifier clairement la session dont la variation est significative, nous appliquons un test t (si les données suivent une distribution normale) ou un test de Mann-Whitney-Wilcoxon.

Enfin, si une variation significative est identifiée pour la troisième session nous vérifions si celle-ci est corrélée à celle de la confiance. Pour cela, nous calculerons le cœfficient de corrélation de Spearman pour les variables $NbAckMeanEvol_{(j)}$ et $TrustRMean_{(j,t)}$.

Résultats :

Nous obtenons les résultats suivant pour le test Shapiro-Wilk avec la variable $NbUDPatrol_{(i,j)}$:

Segment	p-value
1	3,5E-01
2	8,4E-05
3	5,7E-05
4	9,4E-02
5	6,9E-02

Nous utiliserons donc le test Mann-Whitney-Wilcoxon pour comparer deux à deux les différents segments. On obtient le tableau suivant :

Session	1	2	3	4
1	-	-	-	-
2	0,20	-	-	-
3	0,45	0,43	-	-
4	0,12	0,86	0,32	-
5	0,32	0,86	0,74	0,71

De même que pour l'intervention, aucune variation significative n'est observée contrairement à l'évolution de la confiance. A nouveau, on observe que la confiance évolue avec un temps de réponse par rapport à la mise en œuvre de contrôle de l'opérateur. Dans le cadre des acquittements, l'opérateur devient plus attentif au comportement du système et perd peu à peu sa confiance dans le système.

> Sur une échelle de temps courte, l'évolution du degré de confiance et du nombre d'acquittement par focus attentionel n'est plus en opposition. Le degré de confiance évolue avec un temps de réponse au changement du degré attentionel de l'opérateur.

Nous avons vu aux sections 6.5.1 et 6.6 que les variations de confiance et le nombre de contrôle effectué par l'opérateur était corrélés. Nous avons voulu vérifier ce résultat à une échelle de temps plus réduite. En effet, la densité d'évènements de nos scénarios autorise une évaluation de la confiance sur des laps de temps plus courts. Ainsi, au lieu d'une évaluation sur dix minutes (pour les premiers résultats), nous avons effectué notre analyse sur des segments temporels de deux minutes. Nous obtenons un résultat contradictoire. En effet, tandis que la confiance décroît, la quantification des mécanismes de contrôle ne montre aucune variation significative. Nous n'avons plus de corrélation entre ces deux variations. On en conclut alors que la confiance évolue en réponse à la stratégie de contrôle (exprimée au sein du dialogue) de l'opérateur. Autrement dit, lorsque l'opérateur change de stratégie de contrôle, la confiance évolue progressivement vers le niveau de confiance correspondant. Cette évolution, contrairement au changement de stratégie de contrôle, n'est pas instantanée.

Ainsi plus celle-ci se maintient et plus la confiance de l'opérateur diminuera.

6.9 Biais expérimentaux possibles

6.9.1 Effets d'apprentissage : sur l'évaluation

Supposons une première session composée de la réalisation d'une mission (un scénario) et d'une phase de visionnage pour évaluer la confiance, l'entretien au cours

du visionnage va permettre au sujet d'apprendre l'objectif de l'expérimentation. De ce fait au cours des sessions suivantes les résultats seront fortement influencés par cet apprentissage.

On peut supposer que cet apprentissage a pour effet possible :

- de faire intellectualiser au sujet son comportement. Ce dernier n'agira plus de façon instinctive (vis-à-vis de la confiance) mais de manière plus réfléchie.
- de modifier les comportements liés à la confiance. En effet, le sujet, n'agissant plus de manière instinctive, réfléchira au comportement qu'il devrait avoir lorsqu'il fait confiance ou non.

Comme nous ne pouvions disposer de suffisamment de sujet pour créer un groupe témoin, nous avons fait le choix d'une première session où le sujet n'était pas averti contrairement aux suivantes. Nous avons pu voir à la section 6.4 que la connaissance ou non du but expérimental n'avait pas d'incidence significative sur les résultats des évaluations.

6.9.2 Effets d'apprentissage : sur les scénarios

La réalisation de plusieurs sessions avec un même sujet soulève une question sur la constitution des scénarios de chacune des sessions. En effet, afin de pouvoir établir des comparaisons entre chaque session, les scénarios doivent pouvoir être comparés et donc être similaires.

Le défaut d'une trop grande similarité sera l'apparition d'un apprentissage concernant le déroulement du scénario. Par conséquent, le comportement du sujet sera grandement dépendant de ce qui a été appris. C'est pourquoi il faut introduire un aléa suffisant au sein des scénarios.

Pour cela, la plate-forme expérimentale possède un avantage : la génération des alarmes est aléatoire. Ceci signifie que même avec des trajectoires d'intrusions prédéfinies, le scénario ne se déroule pas deux fois de la même manière. Pour un nombre restreint de sessions (1 ou 2) cet aléa est suffisant pour un nombre plus important, il est trop limité.

Afin d'augmenter l'aléa, nous proposons de définir les vecteurs d'entrées des intrus aléatoirement. Les intrus suivront des trajectoires semi-rectilignes [3] vers la cible mais arriveront par des points d'entrées aléatoires. L'aléa devient suffisant pour que le scénario ne soit pas prédictible sur un nombre plus important de session. Mais de par leur similarité, les trajectoires d'intrusions restent donc comparables.

Grâce à notre générateur de scénario, nous avons donc pu établir des scénarios avec une part d'aléa suffisante pour éviter ce phénomène d'apprentissage au cours des sessions successives.

[3] ligne droite segmentée dont les points de raccord inter-segment ont été bruité afin de les écarter légèrement de leur position d'origine (voir description du simulateur section 6.8.3).

6.10 Conclusion

L'objectif de cette expérimentation a été de valider notre modèle de dialogue pour l'évaluation de la confiance. Autrement dit, on veut vérifier que les mécanismes de contrôle mis en avant lors de la définition de notre modèle permettent d'avoir un aperçu du degré de confiance d'un opérateur en interaction avec un système.

Le contrôle tel que défini par Castelfranchi se décompose selon deux aspects :

- Le suivi de la tâche : qui se traduit au sein de notre modèle par les mécanismes de demandes d'extension de l'information, de demandes d'explications et les acquittements. Nous avons observé dans notre expérimentation une corrélation forte entre la variation du nombre d'acquittement et la variation de la confiance. C'est-à-dire un lien fort entre le focus attentionel de l'opérateur et sa confiance. En effet, cette corrélation démontre une plus grande focalisation de l'attention de l'opérateur sur la tâche. Ainsi la mesure de ce premier mécanisme de contrôle peut être utilisée comme indicateur de la confiance.
  Par ailleurs nous avions inséré dans notre modèle de dialogue deux autres mécanismes de contrôle. Les demandes d'informations complémentaires et d'explications n'ont pas suivi une évolution parallèle à la confiance. Nous pensons que notre expérience n'était finalement pas adaptée pour évaluer cette partie de notre modèle. En effet, la tâche à accomplir était assez simple et ne nécessitait pas en cas de difficulté, ou de comportement anormal de la part du système, une analyse profonde de la situation. Autrement dit, lorsque quelque chose ne convenait pas à l'opérateur, il agissait par lui-même sans avoir le besoin d'approfondir sa compréhension de la situation. Si le système semblait ne pas réaliser correctement sa tâche, il était évident pour l'opérateur qu'il y avait un problème. Nous pensons que dans le cadre d'une activité plus complexe (la réussite ou l'échec du système est beaucoup plus ambigu), les mécanismes de contrôle que nous avons introduit auraient eu une utilité réelle pour l'opérateur.
  Ainsi, sur la notion de contrôle en tant que suivi de la tâche notre modèle n'est que partiellement validé.
- l'intervention sur la tâche : qui se traduit par la mise en œuvre de dialogues de la couche programmation/intervention à l'initiative de l'opérateur au sein de notre modèle. Nous avions déjà constaté lors de la précédente expérimentation (chapitre 4) une augmentation des actions correctives de l'opérateur lors de la diminution de sa confiance. Dans cette nouvelle expérience, nous retrouvons à nouveau une corrélation sur la variation du nombre d'interventions de l'opérateur et sa confiance. Il est à noter que ce lien n'est observable que pour les automates dont le fonctionnement est mis en doute par l'opérateur, c'est-à-dire l'interception mais aussi la patrouille.

Enfin, le dernier résultat que nous avons présenté concerne l'évolution de la confiance au sein de la troisième session. En effet, l'analyse de l'évolution de la confiance intra-session et des mécanismes de contrôle de notre modèle qui ont été

validés montre que l'évolution de la confiance est progressive contrairement aux mécanismes de contrôle. Nous avons donc un temps de réponse beaucoup plus important pour la confiance, que pour la stratégie de dialogue de l'opérateur. Cela signifie que les corrélations observées ne sont applicable qu'avec des échelles de temps suffisamment large pour que le temps de réponse de la confiance soit négligeable. En revanche cela démontre que notre modèle nous permet donc d'anticiper la dégradation de la confiance de l'opérateur.

6.11 Discussion

Pour rappel, nous basons nos travaux et nos analyses sur l'hypothèse que confiance et contrôle sont liés par une relation d'opposition : le contrôle se substitue à la confiance lorsque celle-ci devient insuffisante [35]. Cette hypothèse repose sur le fait que Homme et machine établissent une relation de coopération dans le cadre du contrôle supervisé. C'est pourquoi nous avons décidé d'évaluer la confiance à partir de l'observation des contrôles d'un opérateur sur son système.

L'expérience nous a bien montré qu'un opérateur plus attentif à son système exprimait, aux travers de l'évaluation de la confiance par questionnaire, une confiance moindre. Par contre, la possibilité donnée à l'opérateur d'interagir avec le système pour vérifier plus finement le comporte de son système n'a pas été exploitée. Si l'on se réfère au modèle de la confiance de Lee [66], la confiance d'un opérateur dépend en partie de l'historique des succès et/ou des échecs des tâches accomplies par le système mais aussi de sa représentation mentale du système. Les mécanismes d'explication et d'extension de l'information sont censés aider l'opérateur à établir une représentation du système. Or ils ont été délaissés par les sujets qui se sont contentés de l'observation des succès ou échecs du système mais ce serait oublier la faible complexité de notre simulateur. En effet, celui-ci ne nécessitait peut-être pas une représentation mentale approfondie pour sa bonne supervision. Ainsi bien que ces mécanismes n'aient pas été mis en relation avec la confiance nous maintenons leur utilité au sein de notre modèle.

Le deuxième point de discussion est l'emploi de la notion de grounding dans notre modèle. La théorie qui l'accompagne et qui est utilisée comme point de départ de notre propre modèle décrit les mécanismes de partage de l'information au sein d'un dialogue. On constate au travers de notre expérience que l'opérateur fait plus attention aux informations partagées par le système lorsque la confiance décroit. Par ailleurs, on a constaté que les interventions de l'opérateur étaient ciblées, et concernaient principalement la tâche d'intervention. On se rend compte alors que le contenu sémantique du dialogue a aussi son importance. En effet, la sémantique nous renseigne sur l'objet de confiance et donc sur la fonctionnalité du système dont l'opérateur est méfiant.

Un aspect important du grounding n'a pas été exploité : le degré de grounding. Dans un dialogue qui passe par une interface graphique les variations dans l'expression d'un acquittement sont très limitées si ce n'est unique. Il n'est donc pas

possible d'appliquer l'échelle de grounding définie par Roque [96]. Il serait tout de même intéressant d'approfondir ce point en mettant en œuvre des interactions qui permettraient une expression de plusieurs degrés de grounding différents.

Enfin, notre modèle ne permet pas de décrire de manière précise l'évolution de la confiance. En effet, nous avons vu qu'à différentes échelles de temps nous n'obtenions pas les mêmes résultats. Ainsi sur des échelles de temps courtes, nous avons observé un changement radical dans le comportement de l'opérateur tandis que la confiance n'évoluait que progressivement. Nous avons avancé l'idée que l'opérateur réévaluait progressivement sa confiance au fur et à mesure qu'il maintenait une stratégie d'interaction axée sur un contrôle fort du système. Or si l'on considère que cette situation de méfiance est due à une panne de l'automate d'interception, il est aussi tout à fait normal que l'opérateur change rapidement sa façon d'interagir. Mais la question qui se pose est de savoir si une perte de confiance avec un automate parfaitement fonctionnel induirait un tel changement comportemental. On constate, tout de même, au première abord que le lien entre interaction et confiance existe et est mesurable. Nous en proposons une description, a priori suffisante, mais limitée.

Conclusion et perspectives

Nous nous sommes intéressés au sein de ce mémoire à l'évaluation de la confiance dans le cadre du contrôle supervisé. Ce domaine met en interaction des systèmes complexes avec des humains dont la confiance envers ces automates n'est pas sans conséquence. Nous avons pu voir au sein de notre état de l'art sur le contrôle supervisé que la confiance pouvait conduire à un mauvais usage (voire un rejet) d'un système autonome. Ces situations aboutissent à des contre-performances du couple Homme-machine. C'est pourquoi, afin d'anticiper ces situations nous avons proposé d'évaluer la confiance des opérateurs.

La confiance est une notion subjective qui dépend essentiellement des perceptions. De ce fait on ne trouve au sein de la littérature que très peu de proposition de mesure objective de la confiance. En effet nous avons pu voir que l'évaluation de la confiance est principalement réalisée à partir de questionnaire a posteriori — trop tard pour anticiper un mauvais usage d'un automate. Nous avons alors proposé une nouvelle approche basée sur l'interaction et l'analyse du dialogue.

Le dialogue est un élément central du contrôle supervisé. En effet le contrôle supervisé implique une interaction continue entre l'homme et la machine. De plus cette interaction, lorsqu'elle est à l'initiative de l'opérateur, n'est que l'expression des intentions de celui-ci. En d'autres mots, la confiance influence en partie les stratégies d'interaction de l'opérateur.

Nous avons mis en évidence le lien entre confiance et stratégie d'interaction lors d'une première campagne expérimentale. Mais cette campagne nous a montré qu'il était nécessaire d'approfondir l'analyse de l'interaction si l'on voulait aboutir à un modèle de l'évaluation de la confiance basée sur l'analyse du dialogue.

En considérant le lien entre les notions de confiance et de contrôle comme une relation d'opposition, nous avons alors défini notre propre modèle de dialogue afin de mettre en avant la notion de contrôle. Par ailleurs nous pensons que la théorie du grounding, comme modélisation du dialogue, reflète une certaine forme de contrôle sur l'information. C'est pourquoi nous avons adapté le modèle de grounding de Traum [108] à différentes couches du contrôle supervisé (suivi et intervention). La quantification des mécanismes de contrôle sert alors de base pour modéliser l'évaluation de la confiance.

Afin de valider notre modèle d'évaluation, une campagne expérimentale a été réalisée. Celle-ci nous a permis de montrer que l'hypothèse d'un lien entre contrôle et confiance était valable. De plus, on a pu montrer que notre modèle de dialogue soutenait cette approche en permettant une mesure des mécanismes de contrôle au cours du dialogue. Mais nous n'avons pu valider l'ensemble de ces mécanismes. En effet les mécanismes d'explication et d'extension de l'information ne peuvent être évalués au travers de cette expérience. Il est donc nécessaire de renouveler

l'expérience avec un système suffisamment complexe pour rendre utile l'usage de ces mécanismes. On pourrait alors étudier leur usage et vérifier que cette forme de contrôle est aussi liée aux évolutions de la confiance.

Enfin notre étude nous permet de poser les bases d'un outil d'évaluation de la confiance. Celui-ci nous permet d'anticiper sur les situations où la confiance se dégraderait dangereusement (risque d'un mauvais usage de l'automate), mais ne résout pas le problème pour autant. Il est donc maintenant nécessaire d'étudier comment au travers d'une interaction adaptative de l'automate il est possible de reconstruire la confiance de l'opérateur.

Bibliographie

[1] Principes de conception de l'interface homme-machine. In *EPR Rapport préliminaire de sureté de Flamanville 3*, chapter 17.3. Edf edition, Juillet 2006.

[2] La france envoie des robots pour intervenir dans la centrale de fukushima. *Le Point.fr*, mars 2011.

[3] Eugenio Alberdi, Lorenzo Strigini, Andrey A Povyakalo, and Peter Ayton. Why are people's decisions sometimes worse with computer support ? *Proceedings of the 28th International Conference onComputer Safety, reliability and security*, pages 18–31, 2009.

[4] Harald Aust and Martin Oerder. Dialogue control in automatic inquiry systems. In *ESCA Workshop on spoken dialogue systems*, 1995.

[5] J L Austin. *How to do things with words*. Oxford university press, 1962.

[6] B. Barber. *The Logic and Limits of Trust*. Rutgers University Press, 1983.

[7] R Bhattacharya, T M Devinney, and M M Pillutla. A formal model of trust based on outcomes. *The Academy of Management Review*, 23(3) :459–472, 1998.

[8] Katinka M Bijlsma and Gerhard G Van De Bunt. Antecedents of trust in managers : a bottom up approach. *Personnel Review*, 32(5) :638–664, 2003.

[9] A. M. Bisantz and Y. Seong. assessment of operator trust in and utilization of automated decision-aids under different framing conditions. *International Journal of Industrial Ergonomics*, 28(2) :85–97, 2001.

[10] Elizabeth K Bowman. Human trust in networks. In *14th International Command and Control Research and Technology Symposium*, 2009.

[11] John R Boyd. The essence of winning and losing. 1996.

[12] S. Bruni, J. J. Marquez, A. Brzezinski, C. Nehme, and Y. Boussemart. Introducing a human-automation collaboration taxonomy (hact) in command and control decision-support systems. In *12th International Command and Control Research and Technology Symposium*, 2007.

[13] C. Castelfranchi and R. Falcone. Social trust : cognitive anatomy, social importance, quantification and dynamics. *Autonomous Agents '98 Workshop on "Deception, Fraud and Trust in Agent Societes"*, pages 35–49, 1998.

[14] C. Castelfranchi and R. Falcone. The dynamics of trust : from beliefs to action. *Autonomous Agents '99 workshop on "Deception, Fraud and Trust in Agent Societes"*, 1999.

[15] C. Castelfranchi and R. Falcone. Trust is much more than subjective probability : mental components and sources of trust. *Proceedings of the 33rd Hawaii International Conference on System Sciences*, 2000.

[16] Cristiano Castelfranchi and Rino Falcone. Trust and control : a dialectic link. *Applied Artificial Intelligence*, 14 :799–823, 2000.

[17] Jessie Y C Chen. Concurrent performance of military tasks and robotics tasks : effects of automation unreliability and individual differences. In *Proceeding of HRI'09*, 2009.

[18] Jessie Y C Chen, Michael J Barnes, and Caitlin Kenny. Effects of unreliable automation and individual differences on supervisory control of multiple ground robots. In *Proceeding of HRI'11*, 2011.

[19] M Cherubini, J Van Der Pol, and P Dillenbourg. Grounding is not shared understanding : Distinguishing grounding at an utterance and knowledge level. *CONTEXT'05, the Fifth International and Interdisciplinars Conference on Modeling and Using Context*, 2005.

[20] Jennifer Chu-Carroll and Sandra Carberry. Response generation in collaborative negotiation. *ACL '95 Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, 1995.

[21] Gavin E Churcher, Eric S Atwell, and Clive Souter. Dialogue management systems : a survey and overview. Technical report, University of Leeds, February 1997.

[22] Herbert H Clark. *Using language*. Cambridge university press, 1996.

[23] Herbert H Clark and Edward F Schaefer. Contributing to discourse. *Cognitive Science*, 1989.

[24] M. S. Cohen, R. Parasuraman, and J. T. Freeman. Trust in decision aids : a model and its training implications. *Proceedings 1998 Command and Control Research and Technology Symposium*, pages 1–37, 1998.

[25] P R Cohen and C R Perrault. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3, 1979.

[26] Philip R Cohen. *Discours and Dialogue : Dialogue Modeling*, chapter 6.3, pages 204–210. Cambridge University Press, 1997.

[27] Philip R Cohen and Hector J Levesque. Teamwork. In *Noûs*, volume 25 of *Special issue on cognitive science and artificial intelligence*, 1991.

[28] Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten Van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction*, 18, 2008.

[29] Henriette S M Cramer, Vanessa Evers, Maarten W Van Someren, and Bob J Wielinga. Awareness, training and trust in interaction with adaptive spam

filters. In *Proceedings of the 27th international conference on human factors in computing systems*, pages 909–912, April 2009.

[30] J. W. Crandall and M. L. Cummings. Attention allocation efficiency in human-uv teams. In *AIAA infotech@Aerospace Conference*, Rohnert Park, May 2007.

[31] Eric A Cring and Adam G Lenfestey. Architecting human operator trust in automation to improve system effectiveness in multiple unmanned aerial vehicle (uav) control, 2009.

[32] M. L. Cummings, S. Bruni, S. Mercier, and P. J. Mitchell. Automation architecture for single operator, multiple uav command and control. *The International Command and Control Journal*, 1(2), 2007.

[33] M L Cummings and Stephanie Guerlain. Developing operator capacity estimates for supervisory control of autonomous vehicles. *the journal of the human factors and ergonomics*, 49, 2007.

[34] M. L. Cummings and P. J. Mitchell. predicting controller capacity in supervisory control of multiple uavs. *IEEE Transactions on Systems, Man, and Cybernetics*, 38(2) :451–460, 2008.

[35] Steven C Currall and Timothy A Judge. Measuring trust between organizational boundary role persons. *organizational behavior and human decision processes*, 64, 1995.

[36] T K Das and B-S Teng. Trust, control and risk in strategic alliances : an integrated framework. *organization studies*, 22 :251–283, 2001.

[37] T K Das and Bing-Sheng Teng. Between trust and control : Developing confidence in partner cooperation in alliances. *The Academy of Management Review*, 23(3) :491–512, 1998.

[38] A K Dasgupta and D W Pearce. *Cost-Benefit Analysis : Theory and Practice*. MacMillan, 1972.

[39] Henri C Dekker. Control of inter-organizational relationships : evidence on appropriation concerns and coordination requirements. *Accounting, Organizations and Society*, 29(1) :27 – 49, 2004.

[40] M Deutsch. Trust and suspicion : Theorical notes. In *The Resolution of Conflict : Constructive and Destructive Processes*, chapter Part One : Theorical Essays, pages 143–176. Yale university press edition, 1973.

[41] Stephen R Dixon, Christopher D Wickens, and Dervon Chang. Mission control of multiple unmanned aerial vehicles : a workload analysis. *the journal of human factors and ergonomics*, 47, 2005.

[42] J. Drury and S. Scott. Awareness in unmanned aerial vehicle operations. *The International C2 Journal*, 2008.

[43] M T Dzindolet, S A Peterson, R A Pomranky, L G Pierce, and H P Beck. The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 58(6) :697–718, 2003.

[44] M. T. Dzindolet, L. G. Pierce, H. P. Beck, L. A. Dawe, and B. W. Anderson. predicting misuse and disuse of combat identification systems. *Military Psychology*, 13(3) :147–164, 2001.

[45] M. R. Endsley and D. J. Garland. *Situation awareness : analysis and measurement*. Lawrence Erlbaum Associates, 2000.

[46] Annika Flycht-Eriksson. A survey of knowledge sources in dialogue systems. In *Proceedings of IJCAI'99 workshop on Knowledge and Reasoning in Practical Dialogue Systems*, pages 41–48, August 1999.

[47] Amos Freedy, Ewart DeVisser, and Gershon Weltman. Measurement of trust in huamn-robot collaboration. In *international symposium on collaborative technologies and systems*, volume 1, 2007.

[48] D. Gambetta. *Trust : Making and Breaking Cooperative Relations*. Basil Blackwell, 1988.

[49] J. Gao and J. D. Lee. Extending the decision field theory to model operators' reliance on automation in supervisory control situations. *IEEE transactions on systems, man and cybernetics. Part A, Systems and humans*, 36(5) :943–959, 2006.

[50] R. Hess and B. D. McNally. Automation effects in a multi-loop control system. *IEEE transactions on systems, man and cybernetics*, SMC-16(1) :111–121, 1986.

[51] B. Hilburn, P. G. A. M. Jorna, and R. Parasuraman. The effect of advanced atc automation on mental workload and monitoring performance - an empirical investigation in dutch airspace. In *8th International Symposium on Aviation Psychology*, pages 387–391, 1995.

[52] J Hulstijn, R Streetskamp, H ter Doest, S van de Surgt, and A Nijholt. Topics in schisma dialogues. In *Dialogue Management in Natural Language Systems*, 1996.

[53] Gail Jefferson. Side sequences. In D Sudnow, editor, *Studies in social interaction*. Free press, 1972.

[54] J-Y Jian, A M Bisantz, and C G Drury. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics,*, 4(1) :53–71, 2000.

[55] Xiaochun Jiang, Mohammad T Khasawneh, Reena Master, Shannon R Bowling, Anand K Gramopadhye, Brian J Melloy, and Larry Grimes. Measurement of human trust in a hybrid inspection system based on signal detection theory measures. *International journal of industrial ergonomics*, 34, 2004.

[56] G. Johannsen. Design of intelligent human-machine interfaces. In *IEEE International Workshop on Robot and Human Communication*, 1994.

[57] G. Johannsen. Knowledge-based design of human-machine interfaces. In *Control Engineering Practice*, volume 3, pages 267–273, 1995.

[58] G. Johannsen and E. A. Averbukh. Human performance models in control. In *Systems, Man and Cybernetics. International Conference on Systems Engineering in the Service of Humans*, volume 4, 1993.

[59] G. Johannsen, A. H. Levis, and H. G. Stassen. Theorical problems in man-machine systems and their experimental validation. *Automatica (Journal of IFAC)*, 30 :217–231, 1994.

[60] D. Kaber, E. Omal, and M. Endsley. Levels of automation effects on telerobot performance and human operator situation awareness and subjective workload. In *Technology and Human Performance : Current Research and Trends*, pages 165–170, 1999.

[61] B H Kantowitz. Pilot workload and flightdeck automation. In Mustapha Parasuraman, R ;Mouloua, editor, *Automatisation and human performance : heory and applications*. Routledge, 1996.

[62] B. H. Kantowitz, R. J. Hanowski, and S. C. Kantowitz. Driver acceptance of unreliable traffic information in familiar and unfamiliar settings. *Human Factors*, 39 :164–176, 1997.

[63] E. Langer. *Mindfulness*. Addison Wesley Publishing Company, 1989.

[64] J. Lee and N. Moray. Trust, control strategies and allocation of function in human-machine systems. *ergonomics*, 35(10) :1243–1270, 1992.

[65] J. D. Lee and N. Moray. Trust, self-confidence, and operators' adaptation to automation. *International Journal Human-Computer Studies*, 40 :153–184, 1994.

[66] J D Lee and K A See. Trust in automation : Designing for appropriate reliance. *Human Factors*, 46 :50–80, 2004.

[67] F. Legras. Etude de l'art du partage d'autorité humain-multi-robot. Technical report, ENST Bretagne, 2007.

[68] Olivier Lemon, Alexander Gruenstein, Alexis Battle, and Stanley Peters. Multi-tasking and collaborative activities in dialogue systems. In *SIGDIAL '02 Proceedings of the 3rd SIGdial workshop on Discourse and dialogue*, volume 2, 2002.

[69] N Luhmann. *Trust and Power*. Wiley, 1979.

[70] N. Luhmann and L. Quéré (traducteur). Confiance et familiarité. *Réseaux*, (108) :15–35, 2001.

[71] S. P. Marsh. *Formalising Trust as a Computational Concept*. PhD thesis, University of Stirling, Department of Computing Science and Mathematics, 1994.

[72] Roger C Mayer, James H Davis, and F David Schoorman. An integrative model of organizational trust. *Academy of management review*, 20(3), 1995.

[73] Daniel J McAllister. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38(1), 1995.

[74] Stephane Mercier. *Contrôle du partage de l'autorité dans un système d'agents hétérogènes*. PhD thesis, Institut supérieur de l'aéronautique et de l'espace, 2011.

[75] Marvin Minsky. Frames. In *the society of mind*, 1988.

[76] M. Morphew and C. Wickens. Pilot performance and workload using traffic, displays to support free flight. In *Proceedings of the 42nd Annual Human Factors and Ergonomics Society Conference*, 1998.

[77] K. Mosier and L. J. Skitka. Human decisions makers and automated decision aids : Made for each other ? In M. Parasuraman, R. ; Mouloua, editor, *Automation and Human Performance : Theory and Applications*, pages 210–220. Hillsdale edition, 1996.

[78] K. Mosier, L. J. Skitka, M. Burdick, and S. Heers. Automation bias, accountability, verification behavior. In *Proceedings of the Human Factor and Ergonomics Society 40th Annual Meeting*, pages 204–208, 1996.

[79] A.-I. Mouaddib. Controlling and sharing authority in a multi-robot system. In *Proceedings of the first conference on Humans Operating Unmanned Systems (HUMOUS'08)*, Brest, France, 3-4 Sept. 2008.

[80] B M Muir. Trust in automation : Part i. theorical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37(11) :1905–1922, 1994.

[81] Bonnie M Muir and Neville Moray. Trust in automation : Part ii. experimental studies ofr trust and human intervention in a process control simulation. *ergonomics*, 39(3) :429–460, 1996.

[82] C. Nass and K. M. Lee. Does computer-synthesized speech manifest personality ? experimantal tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology*, 7(3) :171–181, 2001.

[83] David G Novick and Karen Ward. Mutual beliefs of multiple conversants : a computational model of collaboration in air traffic control. *AAAI'93 Proceedings of the eleventh national conference on artificial intelligence*, 1993.

[84] S. O'Hara and N. Dwyer. an agent-based approach to decluttering the interfaces of multi-uav command and control systems. In C. M. Shoemaker, D. W. Gage, and G. R. Gerhart, editors, *unmanned systems technology*, volume 6561, page 11. The International society for optical engineering, apr 9-12 2007.

[85] R Parasuraman. Humans and automation : Use, misuse, disuse, abuse. *Human Factors*, 39(2) :230–253, 1997.

[86] R. Parasuraman, R. Molloy, and I. Singh. Performance consequences of information-induced complacency. *International Journal of Aviation Psychology*, 3 :1–23, 1993.

[87] R. Parasuraman, T. Sheridan, and C. Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics, Part A : Systems and Humans*, 30(3), 2000.

[88] R. Parasuraman, T. Sheridan, and C. Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics*, 30, 2008.

[89] R. Parasuraman, T. B. Sheridan, and C. D. Wickens. Situation awareness, mental workload, and trust in automation : viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, 2(2) :140–160, 2008.

[90] L. Quéré. La structure cognitive et normative de la confiance. *Réseaux*, (108) :125–152, 2001.

[91] P. G. Raeth and J. M. Reising. A model of pilot trust and dynamic workload allocation. In *Aerospace and electronics conference, NAECON 1997, Proceedings of the IEEE 1997 National*, volume 1, pages 49–56, 1997.

[92] J. Rasmussen. *Information rocessing and Human-Machine Interaction : An Approach to Cognitive Engineering*. North-Hollan, 1986.

[93] A. Rauschert, C. Meitinger, and A. Schulte. Experimentally discovered operator assistance needs in the guidance of cognitive and cooperative UAVs. In *Proceedings of the first conference on Humans Operating Unmanned Systems (HUMOUS'08)*, Brest, France, 3-4 Sept. 2008.

[94] J K Rempel, J G Holmes, and M P Zanna. Trust in close relationships. *Journal of Personality and Social Psychology*, 49(1) :95–112, 1985.

[95] V. Riley. Operator reliance on automation : theory and data. In M. Parasuraman, R. ; Mouloua, editor, *Automation theory and applications*, pages 19–35. 1996.

[96] A. Roque and D. Traum. Degrees of grounding based on evidence of understanding. 2008.

[97] S Saget, F Legras, and G Coppin. Cooperative interface for a swarm of UAVs. In *Proceedings of the first conference on Humans Operating Unmanned Systems (HUMOUS'08)*, Brest, France, 3-4 Sept. 2008.

[98] P. Satchell. *Cockpit monitoring and alerting systems*. Aldershot, 1993.

[99] F David Schoorman, Roger C Mayer, and James H Davis. An integrative model of organizational trust : past, present, and future. *Academy of Management Review*, 32(2), 2007.

[100] J R Searle. *Speech acts : an essay in the philosophy of language*. Cambridge university press, 1969.

[101] T. Sheridan and W. Verplank. Human and computer control of undersea teleoperators. Technical report, MIT Man-Machine Systems Laboratory, 1978.

[102] T B Sheridan. *Telerobotics, Automation, and Human Supervisory Control*. MIT Press, 1992.

[103] T B Sheridan. Human supervisory control. In W B Sage, A P ;Rouse, editor, *Handbook of systems engineering and management*, chapter 16. John Wiley and sons, 2009.

[104] J M Sinclair and R M Coulthard. *Towards an analysis of discourse : the English used by teachers and pupils*. Oxford University Press, 1975.

[105] Randall D Spain, Ernesto A Bustamante, and James P Bliss. Towards an empirically developd scale for system trust : take two. In *Proceedings of the human factors and ergonomics society*, 2008.

[106] Peter Squire, Greg Trafton, and Raja Parasuraman. Human control of multiple unmanned vehicles : effects of interface type on execution and task switching times. *HRI'06 Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, 2006.

[107] A Sutcliffe. Trust : From cognition to conceptual models and design. *Lecture Notes in Computer Science*, 4001 :3–17, 2006.

[108] David R Traum. *A Computational Theory of Grounding in Natural Language Conversation*. PhD thesis, University of Rochester, 1994.

[109] A. Tversky. *Preference, Belief, and Similarity : Selected Writings*. 2003.

[110] A. Tversky and D. Kahneman. Belief in the law of small numbers. *Psychological Bulletin*, 76 :105–110, 1971.

[111] A. Tversky and D. Kahneman. Judgment under uncertainty : Heuristics and biaises. *Science*, 185 :1124–1131, 1974.

[112] C. Wickens, S. Gordon, and Y. Liu. *An Introduction to Human Factors Engineering*. Longman, New York, 1998.